CHAPTER 5

JUST HOW DO AFFYMETRIX NORMALIZATION METHODS COMPARE? STATISTICS CONTEMPLATE BIOLOGY

Published in part as:

Botha A-M., Z.H. Swanevelder and N.L.V. Lapitan (2010) Transcript profiling of wheat genes expressed during feeding by two different biotypes of *Diuraphis noxia*. *Journal of Environmental Entomology* **39(4):** 1206-1231.

Introduction

In the last decade researchers in biology has embraced high-throughput systems to measure gene expression at a transcriptome level. In these information intensive techniques, like cDNA microarrays and Affymetrix arrays, thousands of data points are screened and analyzed, with even larger amounts of data points being produced during an analysis. As in all technologies there are ambiguities that need to be corrected for, *i.e.* the intensities of hybridized spots do not necessarily reflects the correct level of an expressed transcript mRNA but also contains errors as a result of the technology – like PCR biases, hybridization effects, *etc*. (Kriel & Russell 2005; Harr & Schlötterer 2006; Steinhoff & Vingron 2006). Slide comparisons, biological replications, experimental designs and the identification of differentially expressed transcripts, further complicate these analyses.

High-throughput array analyses attempt to counter these possible ambiguities in three different rectifying steps (Kriel & Russell 2005). An initial image analysis deals with image inaccuracies like background corrections, hybridization artefacts, *etc*. (Kriel & Russell 2005; Steinhoff & Vingron 2006). Data pre-processing and normalization are usually done together and attempts to get rid of technical uncertainties while standardizing (*i.e.* normalizing) different hybridization events (*i.e.* slides) to the same scale that will allow direct comparisons between slides (Steinhoff & Vingron 2006). All these normalization methods utilize the Affymetrix GeneChip (Affymetrix, USA) arrays perfect match (*PM*) and/or mismatch (*MM*) probe pairs in some way (Affymetrix 2002, http://www.affymetrix.com/support/technical/whitepapers/sadd_whitepaper.pdf). In Affymetrix GeneChip arrays, the mRNA target molecules are represented on the array by probe sets consisting of 11-20 probe pairs. Each probe pair contains a *PM* probe to the target and a *MM* probe that has an altered nucleotide at the middle position, *i.e.* nucleotide 13 of the 25 bp length probe (Irizarray *et al.* 2003). The mismatch probes are used to measure non-specific hybridization. Expression levels of the transcript is calculated using the intensities of the hybridized target molecules to the various probe sets (Irizarray *et al.* 2003). After all these corrections, higher-level analyses are used to obtain statistical significant answers to biological questions (Kriel & Russell 2005). The analyses needed at each of these steps have prompted the development of numerous statistical applications (Smyth & Speed 2003; Smyth 2004; see Table 1 of Irizarray *et al.* 2006 for an extensive list; Elo *et al.* 2009).

However, with such a wide variety of statistical techniques suitable for application to a dataset, each with its own biases, assumptions and ambiguities, choosing the best approach that will produce statistical significant, biologically relevant results, can quickly become a nightmare.

Various comparisons of different normalization and pre-processing methods attempt to identify the best approach, but only add to the confusion with conflicting solutions to this problem (Bolstad *et al.* 2003; Irizarray *et al.* 2003; Shedden *et al.* 2005; Harr & Schlötterer 2006; Irizarray *et al.* 2006; Lim *et al.* 2007). Highly controlled dilution and/or spike-in calibration datasets, used in these comparisons, are blamed by some for the conflicting answers (Shedden *et al.* 2005).

As an alternative to approaches followed in all previous literature/studies, we propose to test the feasibility of using more than one normalization method when identifying the statistically significant genes expressed using an Affymetrix gene array set. Our hypothesis is that a subset of the Affymetrix gene identities (geneIDs)/probe sets, ascertained as differentially regulated, will be the same regardless of the normalization method utilized if all the other analyses are kept identical. This subset is therefore normalization-method-bias-independent and reflects more accurately significant differentially regulated transcripts. We've selected five different background correction and normalization methods to investigate their influences on the identification of specific, differentially expressed gene transcripts. The rest of the statistical methods were kept the same to investigate the relevance of these normalization and background correction methods. This allowed us to compare the unique Affymetrix geneIDs/probe sets ascertained as differentially regulated, at three statistical significant cut-offs: $p \leq 0.5, 0.1, 0.001$, and thus allowing us to assess how many times a specific geneID/probe set was identified by the different methods. We've also explored the effectiveness of false discovery rate (FDR) and family-wise type I error rate (FWER) approaches in nullifying any normalization-biases in order to produce more reliable differentially expressed gene identifications. During the final analyses to determine whether a specific gene was differentially expressed, we applied the criterion that a gene must be present at a significant level ($p \leq 0.5, 0.1, 0.001$) in at least four of the five normalization methods to fulfil the significant selection criteria.

Materials and methods

Experimental design

Two experiments were independently conducted by different researchers, with three different institutions used for the labelling, probing and scanning of the slides (*i.e.* the Centre for Proteomic & Genomic Research (CPGR), Cape Town, South Africa; the Microarray Core Lab (Aurora, CA) and the Virginia Bioinformatics Institute Core facility). A fractional design for the comparison analysis

of both the datasets was used (*Figure 5.1*). The two experiments used in this comparative analysis employed 12 and 18 GeneChip Wheat Genome Arrays (http://www.affymetrix.com/support/technical/datasheets/wheat_datasheet.pdf) and are referred to as the 12 and 18 slide experiments.

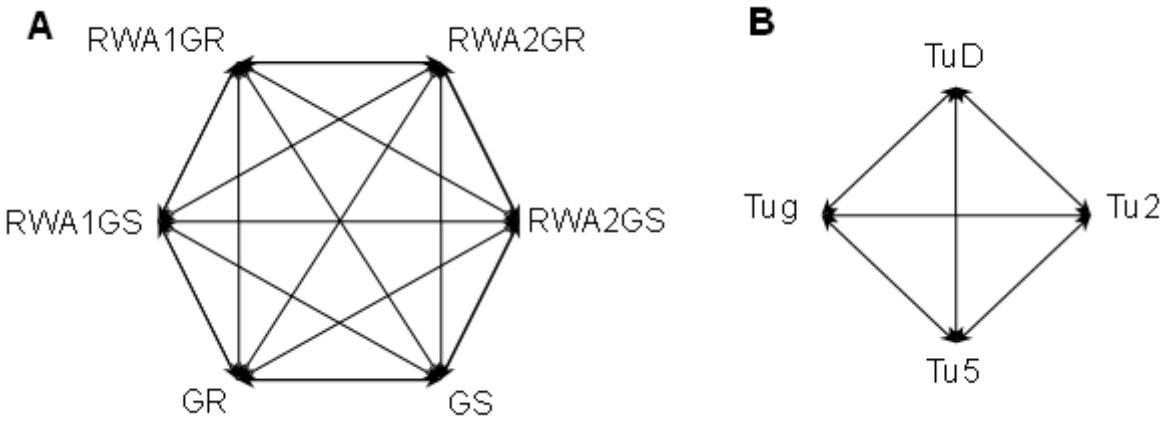

Figure 5.1 A 2 × 2 fractional design used in the analysis of the two experiments, with (A) representing the 18 slide experimental comparison and (B) the 12 slide comparison. Abbreviations (A): RWA1GR, *Dn7* resistant cultivar 94M370 infested with US RWA biotype 1; RWA1GS, the susceptible cultivar Gamtoos infested with US RWA biotype 1; RWA2GR, the *Dn7* resistant cultivar 94M370 infested with the US RWA biotype 2; RWA2GS, susceptible Gamtoos line infested by US RWA biotype 2; GR, the resistant cultivar 94M370 control that was un-infested and GS, the susceptible Gamtoos control, un-infested. Abbreviations (B): Tug, the susceptible cultivar Tugela; TuD, the resistant NIL TugelaDN; Tu2, the resistant NIL Tugela *Dn2* and Tu5, the resistant NIL Tugela *Dn5* - all these lines were infested with the South African RWA biotype.

Aphids and plant material

In the 12 slide experiment, the effects of RWA feeding on near isogenic wheat lines (NILs) between different cultivars infested with one RWA biotype (South African RWA biotype), was investigated, while the 18 slide experiment examined the effects of two RWA biotypes (USA biotypes 1 and 2) on resistant and susceptible wheat lines (Botha *et al.* 2010). In the 12 slide experiment, the South African (SA) RWA biotype (Du Toit 1989), obtained from the ARC-Small Grains Institute (Bethlehem, South Africa) and maintained on the susceptible wheat cultivar Scheepers, was used to infest three resistant (TugelaDN, Tugela *Dn2* and Tugela *Dn5*) lines and one susceptible wheat cultivar (Tugela). In the 18 slide experiment, two USA biotypes, RWA1 - the

original USA 1986 RWA introduction (Lapitan *et al.* 2007) and RWA2 - a biotype virulent to *Dn4* resistant winter wheat cultivars (Haley *et al.* 2004), were used to infest Gamtoos (susceptible) and the resistant line, 94M370 (*Dn7*). In short, in both experiments, plants were grown under greenhouse conditions (20 – 25 °C) until the 4 leaf stage, when plant infestations commenced in accordance to Botha *et al.* (1998). Total RNA was extracted with the Qiagen RNeasy Plant Mini Kit (Qiagen, USA), including on column DNA digestion with RNase Free/DNaseI (Qiagen), all in accordance with the manufacturer's instructions. RNA quality and integrity were tested using agarose gels and the Bio-Rad Experion RNA StdSen Chips (Bio-Rad, USA). The South African RNA samples were sent to CPGR (Cape Town, RSA) and the USA samples to the Microarray Core Lab (Aurora, CA) or to the Virginia Bioinformatics Institute Core facility, where additional quality control was performed. RNA labelling, hybridisation, processing and data gathering were all done in accordance to the Affymetrix protocols at these facilities (Botha *et al.* 2010). A full description of the experimental layout can be found in Botha *et al.* (2010).

Data analysis

Statistical analysis was carried out using R 2.6.0 (http://CRAN.R-project.org/, Ihaka & Gentleman 1996) and Bioconductor 2.1 (http://www.bioconductor.org, Gentleman *et al.* 2004). Additional packages, like the affy (Gautier *et al.* 2004) and affyPLM (Bolstad 2007) packages, were obtained from the Bioconductor website (http://www.bioconductor.org/packages/) as needed. Scripts were executed on a Sun Fire V880 (4 × 1.05 GHz SuperSPARC CPU's with 8 GB RAM) running the Solaris 9 operating system, with all R script outputs written to file. The raw data obtained from the various institutions were analyzed as described in *Figure 5.2* using the scripts, codes and files summarized in *Table Appx 5.1* and script source code (Appendix Chapter 5). In short, information in the raw data files (*.CEL* files), obtained from the various institutions were used for the different pre-processing and summarization methods. Phenotypic data, including biological repeats, technical repeats, sample names to be used in analysis, slide numbers and *CEL* filenames, were stated in a tab delimited text file (*Target.txt*, Appendix Chapter 5) for each of the experiments and included in the raw dataset used for each analysis. The raw data was subjected to different quality control checks which included the inspection of the hybridized images, histograms and box plots of $\log_2$(PM) values and the examination of hybridizations. Background correction and normalization were performed using Affymetrix Microarray Suite5 (MAS5.0) (Harr and Schlötterer, 2006), GeneChip

Robust Multichip Average (GCRMA) (Zakharkin *et al.*, 2005), Probe Level Models (PLM) (Bolstad *et al.*, 2003), Robust Multichip Average (RMA) (Irizarry *et al.*, 2003) and Variance Stabilisation (VSN) (Huber *et al.*, 2002). The outputs of these different methods were saved in corresponding *RData* files (*Figure 5.2*).

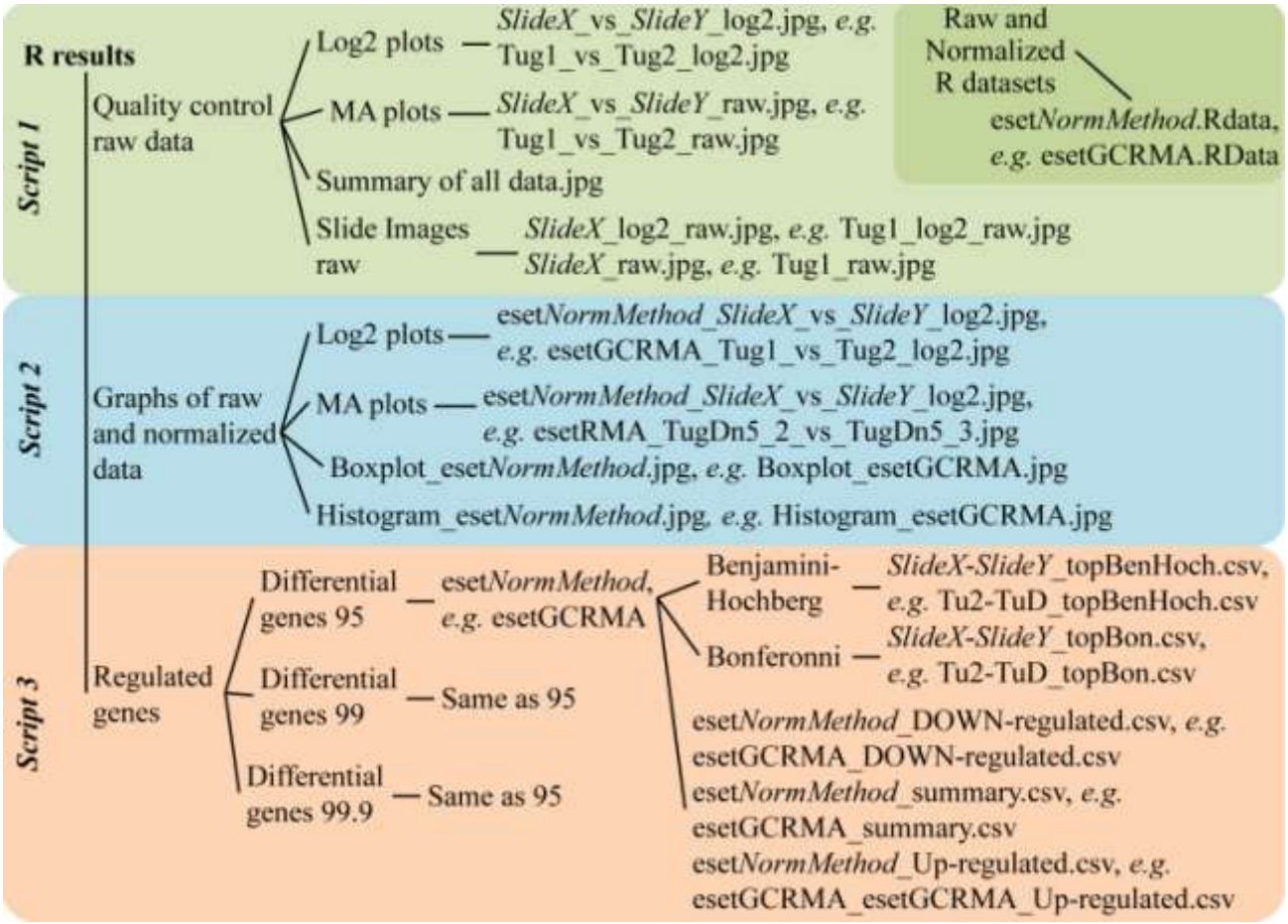

Figure 5.2 A layout of the analysis conducted on each experiment, including some of the output folders and filenames (with examples), as produced by the different R scripts. Files and folders produced during the running of script 1 are represented by a green block, with script 2 and 3 represented by a blue and red block respectively. Variable portions of filenames are given in italics. Details on the input files and the output files/folders are given within the header of each script (*Table Appx 5.1*).

Differentially expressed genes, for datasets of each different background correction and normalization method, were determined for three different *p*-values: 0.05, 0.01 and 0.001. These differentially expressed genes were identified from the different normalized datasets by applying the same moderated *t*-test of the limma (linear models for microarrays, LMM) package and Bayesian analysis (BA) (Smyth & Speed 2003; Smyth 2004). Transcripts were deemed differentially expressed if it had a $\log_2$ fold change larger than 1 ($[\log_2(FC)] >1$) and a *p*-value (or an adjusted p-value) of less-than-or-equal-to 0.05, 0.01 or 0.001 respectively. The comparison-wise false discovery rate (FDR) was controlled with the Benjamini-Hochberg multiple testing adjustment (Benjamini and Hochberg 1995) and the family-wise type I error rate (FWER) by the Bonferroni method (Dudoit & Ge 2004). This produced three datasets before FDR/FWER were determined (95, 99 and 99.9 % confidences) and three for each of the two correction methods.

A MS Excel 2007 macro (*Convert*, Appendix 5) was used to combine all the output spreadsheet files into summary spreadsheets (*Figure 5.3*). The number of times a geneID/probe set was obtained using the different background correction and normalization functions, were calculated in MS Excel 2007. The presence/absence of a regulated GeneID in a slide comparison was represented with a 1/0. Adding these numbers results in the number of times a GeneID/probe set was flagged (for a specific *p*-value) when data was pre-processed differently. The summarized data was used to compile the different graphs using Sigma Plot (Jandel Scientific Inc., USA).

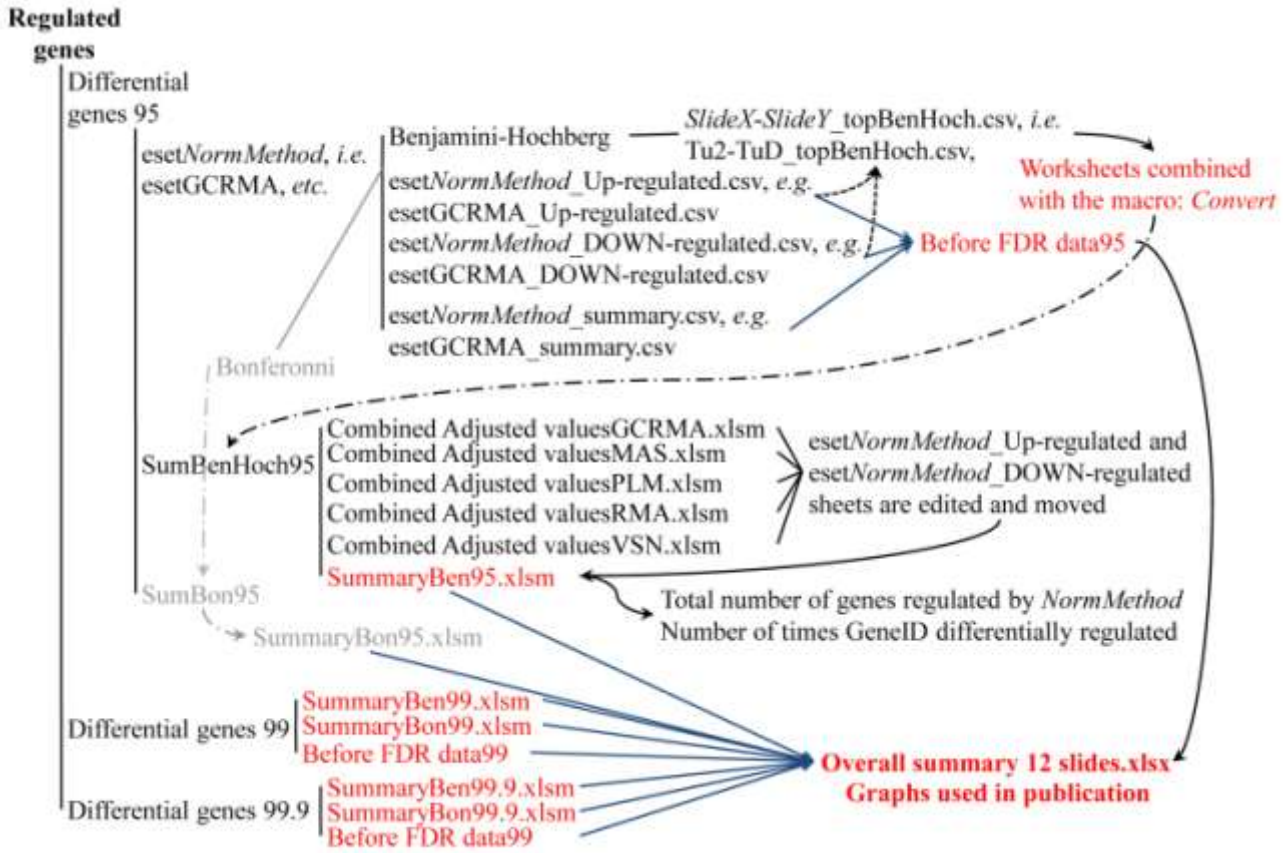

Figure 5.3 The work flow of the data analysis done on the differentially regulated genes, before and after Bonferroni (Bon) and Benjamini-Hochberg (Ben or BenHoch) corrections for 95, 99 and 99.9 % confidences. The same procedures followed for Benjamini-Hochberg analysis were done for Bonferroni. These analyses were repeated for each of the different confidence levels. Files added after the analyses are indicated in red. The summary files are given in bold red.

Results

Quality control of slides, background correction and normalization

Quality control started with an inspection of the original scanned slide images (*Figure Appx 5.1*) to check for any ambiguities on the slides themselves. This was followed by an overall summary figure that contained a histogram, dendogram and box plot of the slides used in the experimental layout (*Figure Appx 5.2*). Other quality control figures produced for the raw data sets, before normalization and background correction, included the $\log_2$ expression comparisons between the pre-normalized slides (*Figure Appx 5.3*) and MA-plots of these slides (*Figure Appx 5.4*). In all

cases where Bioconductor and R modules allowed, additional information like legend names, were incorporated into the figures using the data from the "*Target.txt"* file (Appendix 5). The quality control steps, after normalization and background corrections with GCRMA, MAS5.0, PLM, RMA and VSN, consisted out of summaries of the treatment comparisons using histograms (*Figure Appx 5.5.*), box plots of the $\log_2$(PM) values (*Figure 5.4* and *Figure Appx 5.6*), $\log_2$ expression slide comparisons (*Figure Appx 5.7*) and MA-plot comparisons (*Figure Appx 5.8*). The pre-normalization dataset was also included in these analyses to allow comparisons with the normalized datasets. An example of the results, before and after background correction and normalization with the different methods, are presented in *Figure 5.4* as different box plots of $\log_2$(PM) values of the 12 slide experiment. These box plots, when compared to the raw dataset plot, show the influence of the normalization process by the different methods employed.

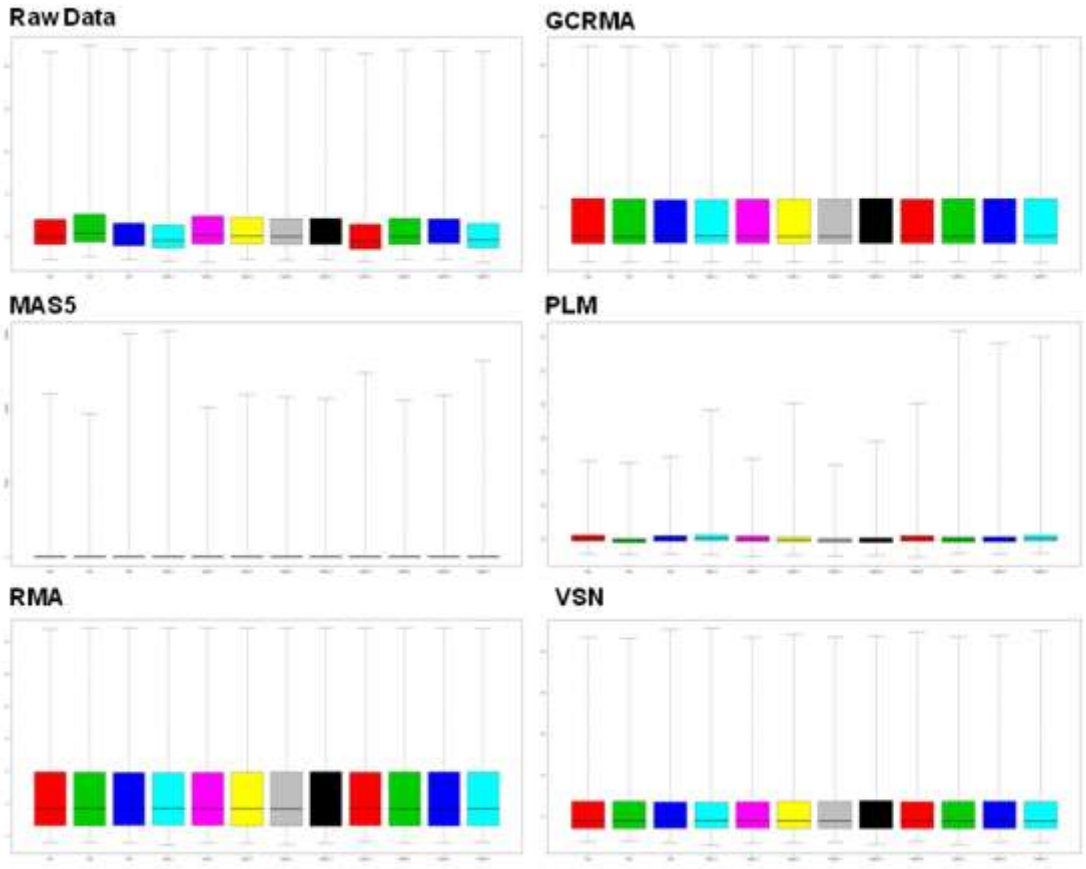

Figure 5.4 Box plots of the $\log_2$ (PM) values, before and after background correction and normalization were done on the 12 slide dataset with the different statistical approaches. All the sample slides occur in the same order in the various graphs in their corresponding triplicate slides (Tugela, TugelaDN, Tugela *Dn2* and Tugela *Dn5*) but that the Y-axis differs in accordance to the method used.

Normalization and background correction methods

The number of times a specific normalization method identified the most or the least number of probe sets, before and after FDR/FWER correction, were compared (*Figure 5.5*). Methods that tend to be less strict in their criteria should more often produce more probe sets than those with stricter criteria. In both the 12 and 18 slide experiments, before FDR/FWER corrections, the PLM method produced the most probe sets across slide comparisons identified as up regulated for all three confidence levels (95, 99 and 99.9 %). The PLM method also produced the most down regulated probe sets more often for the 12 slide experiment slide comparisons, before FDR/FWER correction and across all confidences. The VSN method delivered the most down regulated probe sets for the 18 slide experiment (*Figure 5.5* (*A*) and (*D*)). The MAS5 correction method produced the least number of probe sets more often for both the up and down regulated 12 slide experiment across all the confidence levels and slide comparisons. The VSN method produced more regularly the least number of probe sets for the 18 slide up regulated slide comparisons across the different confidence levels, but different methods contributed for the least number of probe sets for the 18 slide down regulated slide comparisons (*Figure 5.5* (*A*) and (*D*)).

The mean number of probe sets across the difference confidence levels, supported these method percentages (*Figure 5.5*) as producing the largest and smallest number of probe sets for a slide comparison before FDR/FWER correction (*Figure 5.6*). The average probe set number also decreased as the confidence levels increased. However, large differences existed in the number of probe sets identified as differentially regulated for the various slide comparisons within the different experiments. These differences resulted in large errors of their means (*Figure 5.6*). Therefore, significant differences in mean probe set numbers of the different normalization methods are usually only observed between methods producing the mean most and the mean least number of probe sets (*Figure 5.5*). Furthermore, as the number of slides analyzed increased, the average number of probe sets identified as differentially regulated, also increased, from <5 000 in the 12 slide experiment to >10 000 in the 18 slide experiment (*Figure 5*.6), thereby further hampering direct comparisons between the various methods.

In both experiments, the Bonferroni FWER correction and Benjamini-Hochberg FDR correction usually lowered the total mean number of genes identified as differentially regulated (*Figure 5.6*). At the same time these correction methods usually brought the average number of probe sets across different confidence levels, closer together (*Figure 5.6*). In the 12 slide experiment after FWER correction, the RMA method usually produced the most probe sets in the different slide

comparisons (*Figure 5.5* (*B*)). However, for the same experiment, Benjamini-Hochberg FDR correction resulted in the PLM method usually produced the most probe sets under the various normalization conditions (*Figure 5.5* (*C*)). The least number of probe sets after FDR and FWER corrections, for both up and down regulated genes and across all the different confidences, were obtained in 10 of the 12 cases after normalization with the MAS5 method (*Figure 5.5* (*B*) and (*C*)). The VSN method contributed the most genes for the up regulated and the PLM method for the down regulated comparisons, after both FDR and FWER correction and under the various confidence levels for the 18 slide experiment (*Figure 5.5* (*E*) and (*F*)). The Bonferroni FWER corrections for this experiment usually had the MAS5 method producing the least number of probe sets for the different slide comparisons (*Figure 5.5* (*E*)). The least number of probe sets identified as differentially regulated after Benjamini-Hochberg FDR correction was obtained from the MAS5, VSN and the GCRMA methods (*Figure 5.5* (*E*)).

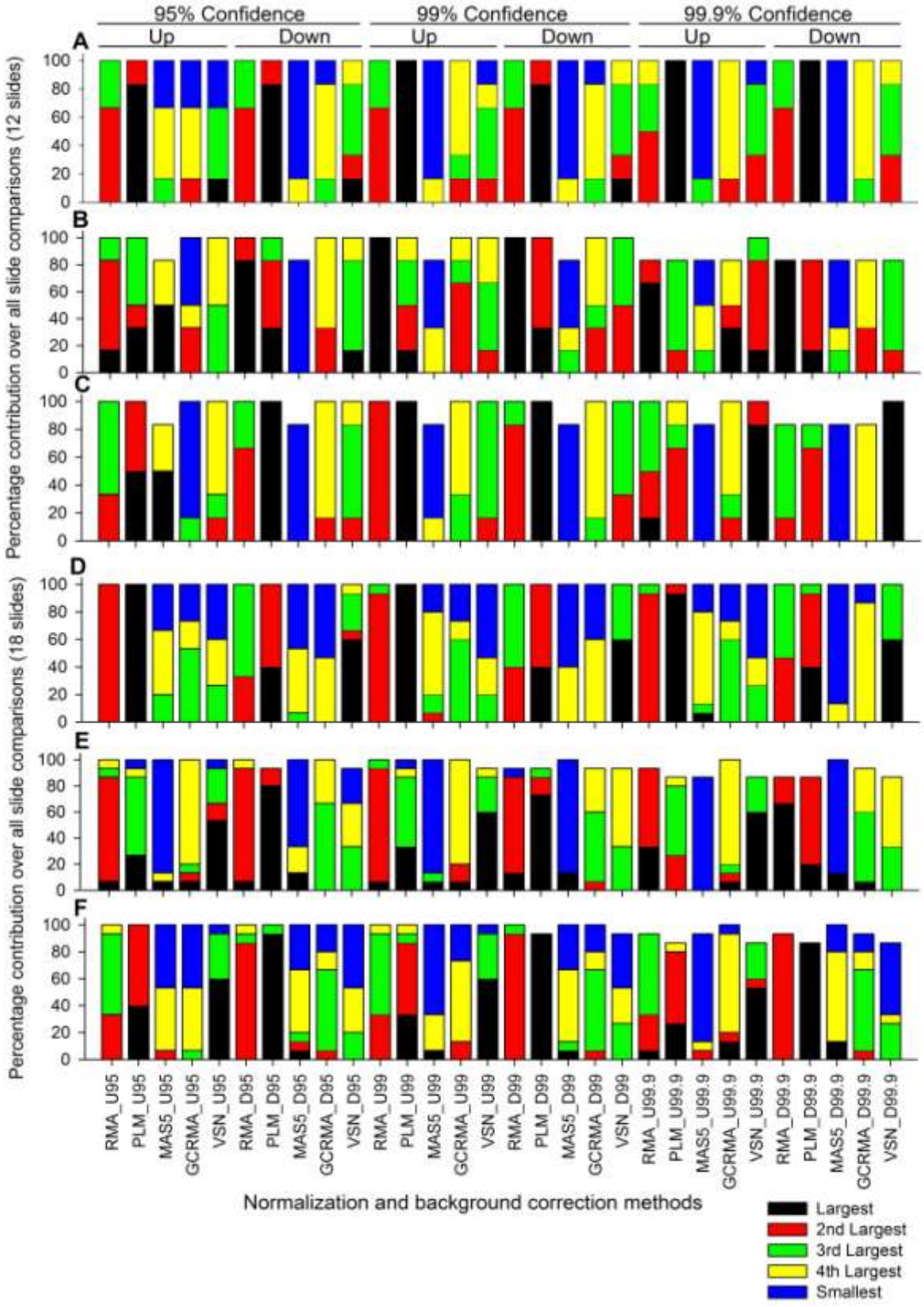
95% Confidence
99% Confidence
99.9% Confidence
Up
Down
A
B
C
D
E
F
Percentage contribution over all slide comparisons (12 slides)
Percentage contribution over all slide comparisons (18 slides)
RMA_U95
PLM_U95
MAS5_U95
GCRMA_U95
VSN_U95
RMA_D95
PLM_D95
MAS5_D95
GCRMA_D95
VSN_D95
RMA_U99
PLM_U99
MAS5_U99
GCRMA_U99
VSN_U99
RMA_D99
PLM_D99
MAS5_D99
GCRMA_D99
VSN_D99
RMA_U99.9
PLM_U99.9
MAS5_U99.9
GCRMA_U99.9
VSN_U99.9
RMA_D99.9
PLM_D99.9
MAS5_D99.9
GCRMA_D99.9
VSN_D99.9
Normalization and background correction methods
Largest
2nd Largest
3rd Largest
4th Largest
Smallest

Figure 5.5 *(previous page)* Normalization method contributions, according to the number of genes differentially up or down regulated, were grouped into the largest, second largest, *etc*. group for each specific slide comparison. The number of times a specific method, expressed as total percentage contribution across all slides, were then calculated and presented here, *i.e.* how many times does a method produce the most, second most, *etc*. number of genes for a specific confidence level, across all the slide comparisons. These values were calculated before FDR/FWER correction ((A), 12 slides and (D), 18 slides) and after FWER/FDR corrections (Bonferroni: (B), 12 slides and (E), 18 slides; Benjamini-Hochberg: (C), 12 slides and (F), 18 slides). The total percentage contribution (y-axis) for that method is less than a 100 % if there are zero genes identified as differentially regulated for one or more slide comparisons. Slide comparisons sometimes delivered the same number of differentially regulated genes for two normalization methods. In these cases, both methods were placed in the same group. Following normalization method identification, "U" indicates up and "D" down regulated comparisons, followed by the confidence level (95, 99 and 99.9 %).

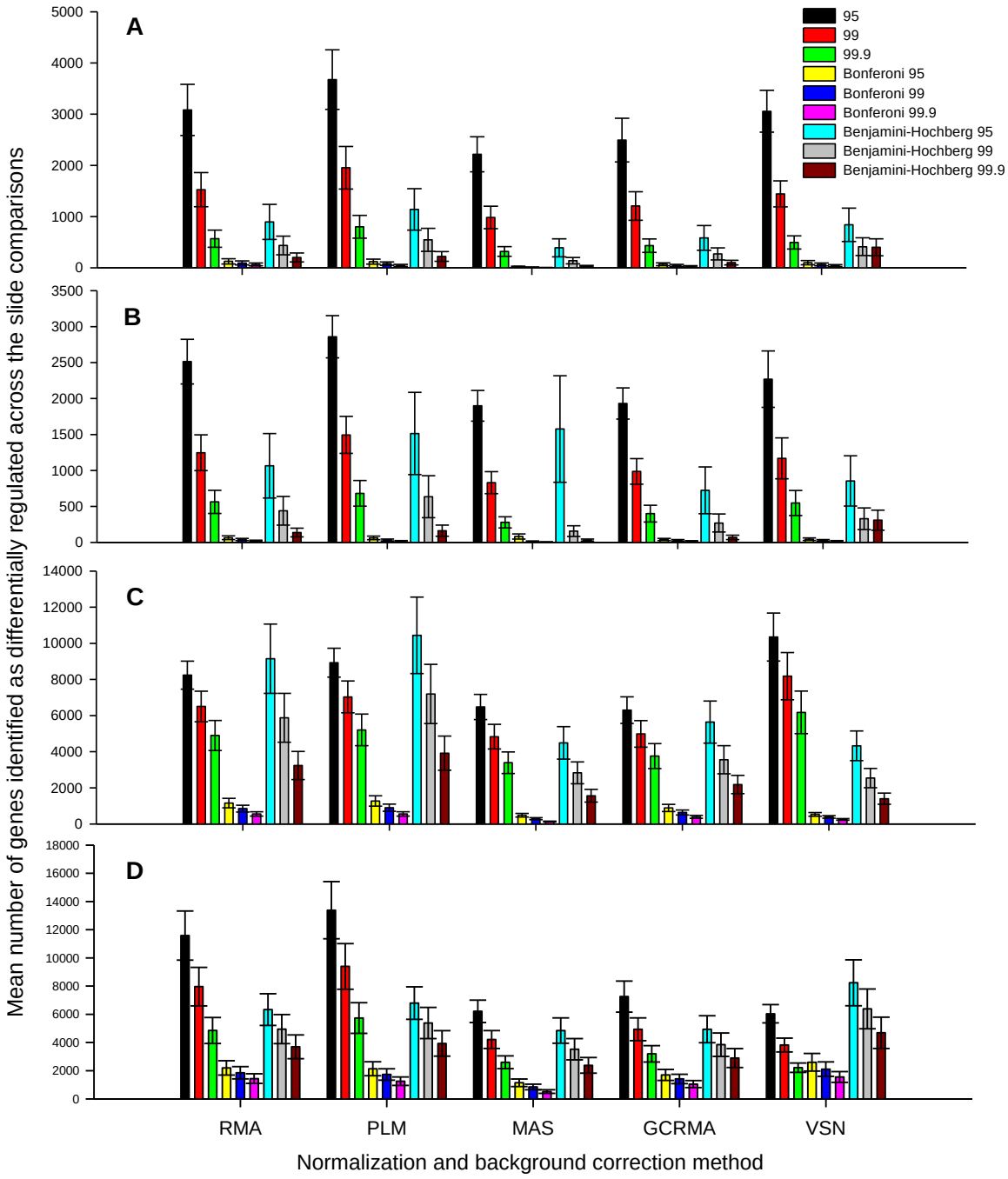

Figure 5.6 The mean number of genes/probe sets found to be differentially regulated as found by the LMM and BA after normalization with the five methods and at 95, 99 and 99.9 % confidence, before/after FDR/FWER corrections. The standard error over all the slide comparisons is indicated by the error bars. (A) and (B) represents the down and up regulated 12 slide experiment respectively, while (C) and (D) represent the down and up regulated genes of the 18 slide experiment, respectively.

Identifying false positives

Differentially expressed genes were identified using linear mixed models (LMM) and Bayesian analysis (BA) for each of the background and normalization methods at 95, 99 and 99.9 % confidence levels and included only those genes that had a $\log_2$ fold change larger than 1. The top differentially regulated probe sets after each normalization method and those unique to a specific slide comparison, for the 12 slide experiment, are given as examples in *Table 5.1* and *Table Appx 5.2*. The FDR/ FWER for these datasets were determined and saved at the set confidence for each slide comparison for a specific background and normalization method. All these datasets were used to calculate the number of times a specific geneID/probe set was differentially regulated after different normalization methods, at different confidences, with/without FWER/FDR correction. Examples of these analyses, *i.e.* with the FDR/FWER corrections for the selected probe sets at each confidence level, are given in *Tables 5.2, 5.3* and *Table Appx 5.3*.

Differentially up- or down regulated genes were determined independently, as well as in combined sets, for both the 12 and 18 slide experiments. These were summarized and used to compute totals of all the percentages of differentially regulated (up, down or combined) GeneIDs/probe sets for the different slide comparisons calculated to be present after 1, 2, 3, 4, or 5 normalization methods, under various confidences (95, 99 and 99.9 %), with/without FDR/FWER corrections. The averages of these total percentages summarize differentially regulated probe sets for all the slide comparisons, thereby allowing for direct comparisons between the methods and experiments (*Figure 5.7*). In all cases, the largest percentages of geneIDs/probe sets were found after normalization by only a single normalization method, regardless of FDR/FWER corrections (*Figure 5.7*). Indeed, an increase in confidence levels and/or with FDR/FWER correction usually increase the percentage probe sets identified as differentially regulated after correction by only a single normalization method, for example a 15 % increase was observed with an increase in confidence (95 vs. 99.9 % confidence, before correction, *Figure 5.7* (C)). Similarly, an average of ca. 30 % at 95 % confidence before FDR/FWER correction can increase to more than 45 % at a 99.9 % confidence with FDR correction (*Figure 5.7* (C)). The number of probe sets deemed differentially regulated after normalization by all 5 methods, as total differentially regulated genes, are < 35 % for the 18 slide experiment, and even smaller for the 12 slide experiment (< 25 %) (*Figure 5.7*).

Table 5.1 Selected geneIDs/probe sets shown to be up regulated (12 slide experiment), either generally or for specific slide comparisons (TuD-Tug, Tu2-Tug, *etc*.). Values under the different normalization methods (VSN, RMA, *etc*.) indicate the number of times the specific probe set was obtained across the six slide comparisons within that specific normalization method, *i.e.* 6 indicates that the probe set was obtained in all six slide comparisons. The slide comparisons indicate the number of times the probe set was obtained after normalization with the five methods, for that specific comparison, *i.e.* a 5 indicates that all five normalization methods detected that probe set for the specific slide comparison. Values summed for the normalization methods equals the values summed for the slide comparisons and are given by the "occurrence" (Occ.) value at specific confidences. *Table Appx 5.2* shows examples of down regulated probe sets.

Probe set ID	***VSN***	***RMA***	***PLM***	***MAS5***	***GCRMA***	***Occ. 95 %***	***TuD - Tug***	***Tu2 - Tug***	***Tu5 - Tug***	***Tu2 - TuD***	***Tu5 - TuD***	***Tu5 - Tu2***	***Occ. 99 %***	***Occ. 99.9 %***
Ta.28669.2.S1_x_at	6	6	6	5	6	**29**	4	5	5	5	5	5	18	16
Ta.9243.1.S1_at	6	6	6	4	6	**28**	4	5	5	4	5	5	21	15
Ta.28669.1.S1_a_at	5	6	6	5	5	**27**	5	5	5	5	5	2	21	18
Ta.10120.1.S1_at	1	1	1	1	1	**5**	5	0	0	0	0	0	3	1
Ta.10120.1.S1_x_at	1	1	1	1	1	**5**	5	0	0	0	0	0	1	0
Ta.10311.1.S1_at	1	1	1	1	1	**5**	5	0	0	0	0	0	4	3
Ta.6995.1.S1_x_at	1	1	1	1	1	**5**	0	5	0	0	0	0	0	0
Ta.1200.1.S1_x_at	1	1	2	2	1	**7**	0	5	0	2	0	0	0	0
Ta.18713.1.S1_s_at	2	1	1	2	1	**7**	0	5	0	2	0	0	1	0
TaAffx.110724.1.S1_at	2	1	1	1	1	**6**	0	0	5	0	0	1	0	0
Ta.12770.1.S1_at	2	2	2	2	2	**10**	0	0	5	0	0	5	2	0
Ta.14005.1.S1_s_at	2	2	2	2	2	**10**	0	0	5	0	0	5	1	0
Ta.10381.1.S1_at	1	1	1	1	1	**5**	0	0	0	5	0	0	1	0
Ta.1055.1.S1_at	1	1	1	1	1	**5**	0	0	0	5	0	0	1	1
Ta.1055.1.S1_x_at	1	1	1	1	1	**5**	0	0	0	5	0	0	2	1
Ta.11414.3.S1_a_at	1	1	1	1	1	**5**	0	0	0	0	5	0	0	0
Ta.6169.1.S1_at	1	1	2	1	1	**6**	0	0	0	0	5	1	1	0
Ta.6626.1.A1_at	1	1	1	2	1	**6**	0	0	0	0	5	1	2	0
Ta.10269.3.S1_at	1	1	1	1	1	**5**	0	0	0	0	0	5	0	0
Ta.10354.2.S1_x_at	1	1	1	1	1	**5**	0	0	0	0	0	5	3	2
Ta.10354.3.S1_x_at	1	1	1	1	1	**5**	0	0	0	0	0	5	2	0

Tug, Tugela; TuD, TugelaDN, Tu2, Tugela Dn2; Tu5, Tugela Dn5

Table 5.2 GeneIDs/probe sets, selected in *Table 5.1*, that were still shown to be differentially regulated after FDR (Benjamini-Hochberg method) or FWER (Bonferroni) correction. Values under the different normalization methods (VSN, RMA, *etc.*) indicate the number of times the specific probe set was obtained across the six slide comparisons (TuD-Tug, Tu2-Tug, *etc.*) within that specific normalization method, *i.e.* 6 indicates that the probe set was obtained in all six slide comparisons. The slide comparisons indicate the number of times the probe set was obtained after normalization with the five methods, for that specific comparison, *i.e.* a 5 indicates that all five normalization methods detected that probe set for the specific slide comparison. Values summed for the normalization methods equals the values summed for the slide comparisons and are given by the "occurrence" (Occ.) value at specific confidences. *Table Appx 5.3* shows examples of differentially up regulated probe sets after FDR/FWER correction.

Probe set ID	***VSN***	***RMA***	***PLM***	***MAS5***	***GCRMA***	***Occ. 95%***	***TuD - Tug***	***Tu2 - Tug***	***Tu5 - Tug***	***Tu2 - TuD***	***Tu5 - TuD***	***Tu5 - Tu2***	***Occ. 99%***	***Occ. 99.9%***
After FDR														
Ta.28669.2.S1_x_at	3	3	5	2	3	**16**	0	4	5	1	5	1	9	6
Ta.9243.1.S1_at	3	3	3	2	2	**13**	0	3	5	0	5	0	8	4
Ta.28669.1.S1_a_at	3	3	4	3	3	**16**	0	5	5	1	5	0	8	4
TaAffx.110724.1.S1_at	1	1	1	0	0	**3**	0	0	3	0	0	0	0	0
Ta.12770.1.S1_at	0	0	1	0	0	**1**	0	0	1	0	0	0	0	0
Ta.14005.1.S1_s_at	0	0	1	0	0	**1**	0	0	0	0	0	1	0	0
Ta.1055.1.S1_at	1	0	0	0	0	**1**	0	0	0	1	0	0	0	0
Ta.1055.1.S1_x_at	1	0	0	0	0	**1**	0	0	0	1	0	0	0	0
Ta.11414.3.S1_a_at	0	0	1	0	0	**1**	0	0	0	0	1	0	0	0
Ta.6169.1.S1_at	0	0	1	0	1	**2**	0	0	0	0	2	0	0	0
Ta.6626.1.A1_at	0	1	1	0	0	**2**	0	0	0	0	2	0	0	0
Ta.10354.2.S1_x_at	1	1	1	0	1	**4**	0	0	0	0	0	4	2	0
Ta.10354.3.S1_x_at	1	1	1	0	0	**3**	0	0	0	0	0	3	1	0
After FWER														
Ta.28669.2.S1_x_at	0	1	1	0	0	**2**	0	0	2	0	0	0	1	0
Ta.9243.1.S1_at	0	1	1	0	0	**2**	0	0	2	0	0	0	1	0
Ta.28669.1.S1_a_at	0	0	1	0	0	**1**	0	0	1	0	0	0	1	0

*Only geneIDs/probe sets of *Table 5.1* were included if still deemed differentially expressed after FDR or FWER correction. Tug, Tugela; TuD, TugelaDN, Tu2, Tugela Dn2; Tu5, Tugela Dn5

Table 5.3 The influences of an increased confidence on the same three geneIDs/probe sets, before and after FDR/FWER correction. Values under the different normalization methods (VSN, RMA, *etc.*) indicate the number of times the specific probe set was obtained across the six slide comparisons (TuD-Tug, Tu2-Tug, *etc.*) within that specific normalization method, *i.e.* 6 indicates that the geneID/probe set was obtained in all six slide comparisons. The slide comparisons indicate the number of times the probe set was obtained after normalization with the five methods, for that specific comparison, *i.e.* a 5 indicates that all five normalization methods detected that probe set for the specific slide comparison. Values summed for the normalization methods equals the values summed for the slide comparisons and are given by the "occurrence" (Occ.) value at specific confidences.

Probe set ID	***Confi-dence***	***VSN***	***RMA***	***PLM***	***MAS5***	***GCRMA***	***Occ.***	***TuD - Tug***	***Tu2 - Tug***	***Tu5 - Tug***	***Tu2 - TuD***	***Tu5 - TuD***	***Tu5 - Tu2***
Before FDR/FWER													
TaAffx.110208.1.S1_at	0.05	6	6	6	6	6	**30**	5	5	5	5	5	5
TaAffx.26346.1.S1_at	0.05	5	6	5	5	6	**27**	5	5	5	5	5	2
Ta.4593.1.A1_at	0.05	6	5	5	5	5	**26**	5	5	5	5	5	1
TaAffx.110208.1.S1_at	0.01	5	6	6	6	6	**29**	5	5	5	4	5	5
TaAffx.26346.1.S1_at	0.01	5	5	5	5	5	**25**	5	5	5	5	5	0
Ta.4593.1.A1_at	0.01	4	4	5	5	4	**22**	5	5	5	2	5	0
TaAffx.110208.1.S1_at	0.001	4	4	6	4	4	**22**	1	5	5	1	5	5
TaAffx.26346.1.S1_at	0.001	3	4	4	4	3	**18**	0	5	5	3	5	0
Ta.4593.1.A1_at	0.001	3	3	3	4	2	**15**	1	5	5	0	4	0
After FDR													
TaAffx.110208.1.S1_at	0.05	3	4	5	3	4	**19**	0	3	5	1	5	5
TaAffx.26346.1.S1_at	0.05	3	3	3	3	2	**14**	0	4	5	0	5	0
Ta.4593.1.A1_at	0.05	3	3	3	3	1	**13**	0	4	5	0	4	0
TaAffx.110208.1.S1_at	0.01	3	3	4	3	3	**16**	0	1	5	0	5	5
TaAffx.26346.1.S1_at	0.01	1	3	2	1	1	**8**	0	1	5	0	2	0
Ta.4593.1.A1_at	0.01	1	1	3	1	1	**7**	0	1	5	0	1	0
TaAffx.110208.1.S1_at	0.001	3	2	3	1	2	**11**	0	0	5	0	4	2
TaAffx.26346.1.S1_at	0.001	1	1	1	0	0	**3**	0	0	3	0	0	0
Ta.4593.1.A1_at	0.001	1	1	1	1	0	**4**	0	0	4	0	0	0

***Table 5.3* cont.**

After FWER													
TaAffx.110208.1.S1_at	0.05	1	1	3	1	1	**7**	0	0	5	0	1	1
TaAffx.26346.1.S1_at	0.05	0	1	0	0	0	**1**	0	0	1	0	0	0
Ta.4593.1.A1_at	0.05	0	0	0	0	0	**0**	0	0	0	0	0	0
TaAffx.110208.1.S1_at	0.01	1	1	2	0	1	**5**	0	0	4	0	1	0
TaAffx.26346.1.S1_at	0.01	0	0	0	0	0	**0**	0	0	0	0	0	0
Ta.4593.1.A1_at	0.01	0	0	0	0	0	**0**	0	0	0	0	0	0
TaAffx.110208.1.S1_at	0.001	0	0	2	0	0	**2**	0	0	1	0	1	0
TaAffx.26346.1.S1_at	0.001	0	0	0	0	0	**0**	0	0	0	0	0	0
Ta.4593.1.A1_at	0.001	0	0	0	0	0	**0**	0	0	0	0	0	0

*For comparison reasons, geneIDs/probe sets no longer deemed differentially expressed after FDR/ FWER correction were included. Tug, Tugela; TuD, TugelaDN, Tu2, Tugela Dn2; Tu5, Tugela Dn5.

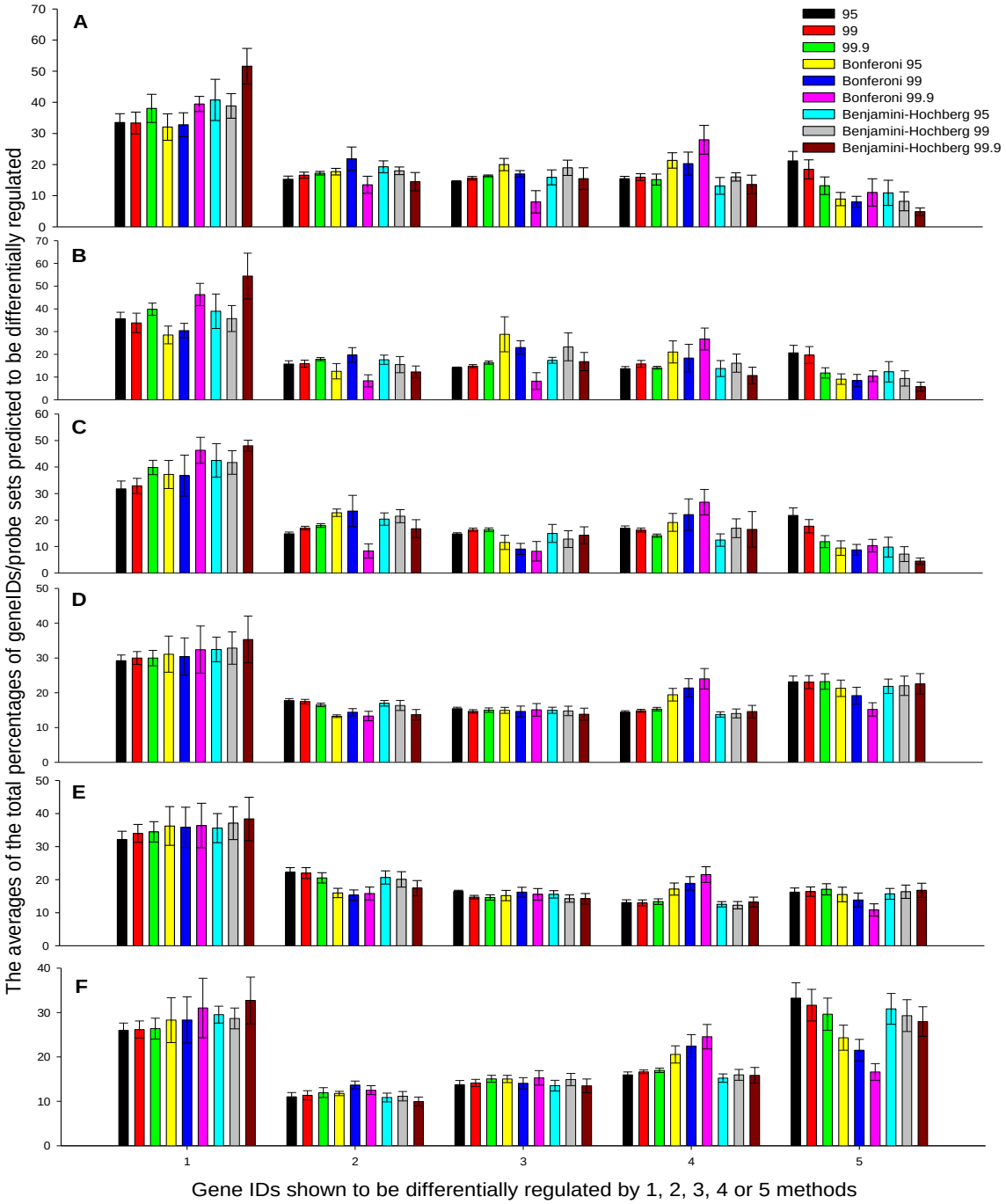

Figure 5.7 The averages of the total percentages of geneIDs/probe sets predicted to be differentially regulated, after normalization at different confidences (95, 99, 99.9 %) and before/after FDR/FWER corrections. The standard errors were calculated on the percentage contribution of each slide comparison to a 1, 2, 3, 4, or 5 method occurrence that was initially used to determine the mean sum (average total) percentage for each specific confidence (95, 99, 99.9 %) and before/after FDR/FWER correction. (A)-(C) represent the 12 slide experiment, with (A) the total differentially regulated genes, *i.e.* the combined down and up regulated percentages, while (B) and (C) represent the down and up regulated genes respectively. Similarly (D)-(F) represent the 18 slide experiment, with (D) the combined, (E) the down and (F) the up regulated differentially expressed gene.

Discussion

Analyzing Affymetrix data can be a daunting prospect, especially with all the numerous normalization and background correction methods available (Smyth & Speed 2003; Smyth 2004; Irizarray *et al.* 2006; Elo *et al.* 2009). The necessity of using these methods is obvious when comparing pre-normalized and normalized datasets, but at the same time the variation in results between the different normalized datasets becomes noticeable (*Figure 5.4* and *Figures Appx 5.2-5.8*). Indeed, normalization procedures are performed to correct for systematic deviations inherent of the technology (Kriel & Russell 2005; Harr & Schlötterer 2006). Increasing the significant cut-offs, *i.e.* confidence levels, or using FDR/FWER corrections methods, can be conceived as suitable countermeasures to any biases introduced by these normalization and background correcting methods. Indeed, an increase in confidence values or in the presence of FDR/FWER correction, the number of times probe sets are categorized as differentially expressed usually does decrease the significant geneIDs/probe sets. However, the same can be observed on a method specific basis (probe sets: Ta.10120.1.S1_at, Ta.10120.1.S1_x_at, Ta.10311.1.S1_at in *Tables 5.1* and *5.2*, also all three accessions of *Table 5.3*). This suggests that an increase in confidence levels or the use of correction methods on a dataset will not neglect the normalization method biases.

The normalization and background correction of Affymetrix slides using different statistical approaches should theoretically produce a large number of similar probe sets, especially if the methods only normalized for the experimental method errors that they were designed for. However, since each method contains inherit biases and statistical assumptions, this is not necessary the case (Kriel & Russell 2005; Shedden *et al.* 2005; Harr & Schlötterer 2006). Indeed, previous studies have shown that normalization methods initially used, have huge implications in detecting differentially regulated genes (Hoffmann *et al.* 2002; Shedden *et al.* 2005), and that these initial correction events have the largest influence on identifying differentially expressed genes (Hoffmann *et al.* 2002; Irizarray *et al.* 2006). This is clearly illustrated in *Figure 5.7*, with the exception for the differentially up regulated part of the 18 slide experiment (*Figure 5.7* (F)), in that more than a third of the geneID/probe set percentages were only associated with a single normalization method. Arguably, this does not necessary indicate that a third of the regulated probe sets are method specific, but that a single normalization method, less strict than the others, could produce more probe sets, a higher probe set percentage and a lower likelihood that another method would have the same probe sets. Similarly, the opposite can also be true. Indeed, there seems to be large differences in the amount of predicted regulated geneIDs/probe sets after normalization with the different

methods, especially at lower confidence levels (95 % No FDR/FWER, *Figure 5.6*). This suggests that an increase in selection criteria stringency, *i.e.* higher confidence levels (99.9 %), fewer geneIDs/probe sets would be unique since highly regulated individuals are preferentially selected. The existence of normalization-method-dependent, differentially regulated, probe set identifications have previously been reported, however only at a single confidence level (Hoffmann *et al.* 2002). Here, the number of probe sets identified as differentially regulated, does decrease to become more comparable when confidence levels increase (99.9 % No FDR/FWER, *Figure 5.6*). However, the averages of the total percentages of geneIDs predicted to be differentially regulated, after normalization at different confidences (95, 99, 99.9 %) and before/after FDR/FWER corrections, occurring in only a single method also increase with this increase in confidence, while those geneIDs/probe sets occurring in all five methods, decrease (*Figure 5.7*). This suggests that a confidence increase does not counteract a normalization method's inherited biases, but does appear to shift geneIDs towards single method identification, *i.e.* increasing the bias.

Normalization method tendencies to produce the most or the least amount of genes for slide comparisons, were investigated for both experiments (*Figure 5.5*). Differences in the number of differentially regulated probe sets have been previously reported for different normalization methods initially used (Hoffmann *et al.* 2002; Shedden *et al.* 2005), as was the case here. In the 12 slide experiment, before FDR/FWER correction, PLM produced the most genes in both up and down regulated slide comparisons at all three confidence levels, with MAS5 the least number of genes (*Figure 5.5*). However, in the 18 slide experiment, PLM only contributed the most genes in the up regulated comparisons, while VSN contributed the largest amount of genes for the down regulated comparisons (*Figure 5.5*). After FDR/FWER correction, however, this switched around, VSN contributes the most genes for the up regulated comparisons, and PLM for the down regulated comparisons. The least number of genes for the up regulated slide comparisons, before FDR/FWER correction, were at each confidence level VSN, while it differed for the down regulated slide comparisons (*Figure 5.5*). FWER correction with Bonferroni of the 18 slide comparisons usually resulted in MAS5 having the smallest number of genes for slide comparisons, while Benjamini-Hochberg FDR corrections having the least number of genes produced by MAS5, VSN and GCRMA (*Figure 5.5*). In the 12 slide experiment, Bonferroni also seemed to lower the number of genes deemed regulated by PLM method, thereby resulting in the RMA method usually producing the most genes in slide comparisons (*Figure 5.5*), but the Benjamini-Hochberg correction did not

(*Figure 5.5*). Again, the MAS5 normalization method usually resulted in the smallest number of gene deemed regulated for 10 out of the 12 overall comparisons (*Figure 5.5*).

False discovery rate (FDR) and FWER correction methods, like Benjamini-Hochberg's and Bonferroni's, tries to lower the number of false positives within predictions of differentially regulated geneIDs/probe sets. From the above results it is clear that normalization methods influence which probe sets will eventually be predicted as differentially expressed (*Figures 5.5* and *5.7*). Furthermore, the probe set eventually identified is dependent on the inherited statistical biases of the normalization method, which may increase as confidence levels increases. These underlying normalization method biases cannot be nullified by FDR/FWER correction methods: Except for Bonferroni at 95 and 99 % confidences, the percentage of probe sets contributed by a single method, increases – especially for Benjamini-Hochberg FDR corrections (*Figure 5.7*). Similarly, the number of probe sets found in all five methods, tend to decrease (*Figure 5.7*). However, Bonferroni FWER correction increased the number of probe sets identified as differentially regulated for the "4 out of 5 normalization methods" class (*Figure 5.7*), though Benjamini-Hochberg's approach didn't when compared with the non-FDR entries. Though the two correction methods tested here usually resulted in a decrease in the number of predicted probe sets (*Figure 5.7*), the same shifting tendency is observed when confidence levels are increased, *i.e.* towards a single geneID-single normalization method, is seen here. The exception, *i.e.* the increased number of probe sets associated with the Bonferroni "4 out of 5 normalization method" class, suggests that this method either decrease false negatives or decrease false positives of a single or multiple methods, thereby increasing the number of probe sets occurring in 4 out of the 5 methods. However, this correction method is very strict and rejects many probe sets, *i.e.* lower geneID numbers (*Figure 5.7*). This suggests that only the highly significant probe sets are retained in each method, thereby increasing the chance of getting the same probe set. A bias in one or more methods probably prevents this FWER correction method from shifting probe sets to the "all 5 methods" class.

New introductions of normalization methods always seem to be better than the last (Irizarray *et al.* 2006). Matters are further complicated by different comparison studies suggesting different normalization method based on different testing approaches; from spike-in data sets (Irizarray *et al.* 2003) to sex linked internal control genes (Galfalvy *et al.* 2003) and co-expressed operon linked control genes (Harr & Schlötterer 2006). Control calibrations/statistics, however, do not consistently produce the same results when using algorithms on experimental datasets (Shedden *et al.* 2005). Indeed, normalization method influence has previous been reported for different study aims, *i.e.*

different methods produced better results for a data set depending on the aim - the identification of differentially expressed genes or co-expressed genes (Harr & Schlötterer 2006; Lim *et al.* 2007). The results presented here suggest that though normalization and background correction methods may produce apparently adequate corrections, they do play a major role in the final geneIDs/probe sets and number of transcripts found to be differentially regulated, as was previously shown (Hoffmann *et al.* 2002; Shedden *et al.* 2005). However, here we further showed that by increasing the confidence levels for selecting differentially regulated probe sets or using FDR/FWER correction methods, the percentage bias of the normalization method initially used may actually increase, while the number of probe sets maybe lowered. How can results be trusted if 20-65 % (*Figure 5.7*) of the geneIDs identified only occurs as the result off one or two normalization methods initially selected? Although probe sets seem to be influenced more by the normalization method initially used, the results highlights the need for using a multiple normalization approach, *i.e.* an approach where the initial normalization and background corrections are done by various methods suitable to the specific experimental design. The various normalized datasets are then carried through the whole process of differentially gene identification, until the resulting probe sets can be compared. This can easily be done using the scripts developed here. If a probe set then occurs in 4 or 5 (or even 3) out of 5 normalized datasets (with/ without FDR/FWER correction), it is more likely to be differentially regulated rather than an introduced normalization method bias. This approach should identify probe sets unique to a specific biological interaction, which is normalization method independent and suitable for marker development.

Acknowledgements

We would like to thank Prof F. Joubert and the Bioinformatics Unit (at the University of Pretoria) for help during the execution of the scripts on their servers, Prof N.L.V. Lapitan (Colorado State University) for collaborating with us on the USA RWA biotypes and Dr. V. Tolmay (Small Grains Institute) for the SA RWA biotype. This research was funded by the Winter Cereal Trust and THRIP (Technology and Human Resources for Industry Programme, South Africa).

References

Affymetrix (2002) Statistical algorithms description document. *Technical report*. Accessed 2009.

Benjamini, Y. and Y. Hochberg (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)* **57**: 289-300.

Bolstad, B. M. (2007) affyPLM: Model Based QC Assessment of Affymetrix GeneChips.

Bolstad, B. M., R. A. Irizarry, M. Astrand and T. P. Speed (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**: 185-193.

Botha, A.-M., M. A. C. Nagel, A. J. van der Westhuizen and F. C. Botha (1998) Chitinase isoenzymes in near-isogenic wheat lines challenged with Russian wheat aphid, exogenous ethylene, and mechanical wounding. *Botanical Bulletin of Academia Sinica* **39**: 99-106.

Botha, A.-M., Z.H. Swanevelder and N.L.V. Lapitan (2010) Transcript profiling of wheat genes expressed during feeding by two different biotypes of *Diuraphis noxia*. *Environmental Entomology: in press.*

Du Toit, F. (1989) Inheritance of resistance in two *Triticum aestivum* lines to Russian wheat aphid (Homoptera: Aphididae). *Journal of Economic Entomology* **82**: 1251-1253.

Dudoit, S. and Y. Ge (2004) Bioconductor's multtest package. *faculty.mssm.edu/gey01/multtest/multtest.pdf*.

Elo, L. L., J. Hiissa, J. Tuimala, A. Kallio, E. Korpelainen and T. Aittokallio (2009) Optimized detection of differential expression in global profiling experiments: case studies in clinical transcriptomic and quantitative proteomic datasets. *Briefings in Bioinformatics* **10**: 547-555.

Galfalvy, H. C., L. Erraji-Benchekroun, P. Smyrniotopoulos, P. Pavlidis, S. P. Ellis, J. J. Mann, E. Sibille and V. Arango (2003) Sex genes for genomic analysis in human brain: internal controls for comparison of probe level data extraction. *BMC Bioinformatics* **4**.

Gautier, L., L. Cope, B. M. Bolstad and R. A. Irizarry (2004) Affy - analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**: 307-315.

Gentleman, R. C., V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dutoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. A. Irizarray, F. Leisch, C. Li, M. Maechler, A. J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L.

Tierney, J. Y. H. Yang and J. Zhang (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology* **5**: R80.

Haley, S. D., F. B. Peairs, C. B. Walker, J. B. Rudolph and T. L. Randolph (2004) Occurrence of a new Russian wheat aphid biotype in Colorado. *Crop Science* **44**: 1589-1592.

Harr, B. and Schlötterer (2006) Comparison of algorithms for the analysis of Affymetrix microarray data as evaluated by co-expression of genes in known operons. *Nucleic Acids Reseach* **34**: e8.

Hoffmann, R., T. Seidl and M. Dugas (2002) Profound effect of normalization on detection of differentially expressed genes in oligonucleotide microarray analysis. *Genome Biology* **3**: research0033.1-0033.11.

Huber, W., A. von Heydebreck, H. Sültmann, A. Poustka and M. Vingron (2002) Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* **18**: s96-s104.

Ihaka, R. and R. Gentleman (1996) R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics* **5**: 299-314.

Irizarray, R. A., B. M. Bolstad, F. Collin, L. M. Cope, B. Hobbs and T. P. Speed (2003) Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Reseach* **31**: e15.

Irizarray, R. A., Z. Wu and H. A. Jaffee (2006) Comparison of Affymetrix GeneChip expression measures. *Bioinformatics* **22**: 789-794.

Kriel, D. P. and R. R. Russell (2005) There is no silver bullet - a guide to low-level data transforms and normalisation methods for microarray data. *Briefings in Bioinformatics* **6**: 86-97.

Lapitan, N. L. V., Y.-C. Li, J. Peng, F. B. Peairs and A.-M. Botha (2007) Limited nuclear and mitochondrial DNA variation among Russian wheat aphid (*Diuraphis noxia*) biotypes from the United Satates and Africa. *Submitted*.

Lim, W. K., K. Wang, C. Lefebvre and A. Califano (2007) Comparative analysis of microarray normalization procedures: effects on reverse engineering gene networks. *Bioinformatics* **23**: i282-i288.

Shedden, K., W. Chen, R. Kuick, D. Ghosh, J. Macdonald, K. R. Cho, T. J. Giordano, S. B. Gruber, E. R. Fearon, J. M. G. Taylor and S. Hanash (2005) Comparison of seven methods for producing Affymetrix expression scores based on False Discovery Rates in disease profiling data. *BMC Bioinfomatics* **6**: 1-12.

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**: Article3.

Smyth, G. K. and T. P. Speed (2003) Limma: linear models for microarray data. Bioinformatics and computational biology solutions using R and Bioconductor. R. Gentleman, V. Carey, S. Dudoit, R. Irizarry and W. Huber. New York, Springer**:** 397-420.

Steinhoff, C. and M. Vingron (2006) Normalization and quantification of differential expression in gene expression microarrays. *Briefings in Bioinformatics* **7**: 166-177.

Zakharkin, S. O., K. Kim, T. Mehta, L. Chen, S. Barnes, K. E. Scheirer, R. S. Parrish, D. B. Allison and G. P. Page (2005) Sources of variation in Affymetrix microarray experiments. *BMC Bioinfomatics* **6**: 1-11.

CHAPTER 6

CONCLUSION

The aphid-plant interaction, especially between *Diuraphis noxia* and wheat, has been the subject of numerous studies. In these, the focus was usually on plant defense: either how different wheat cultivars' respond to RWA feeding or how different aphid biotypes infestation profiles differ based on resistant plant differentials. The theory that resistant cultivars forced new RWA biotypes, *a.k.a.* the arms race, is widely accepted. However, aphids' successes as plant sap feeders are directly contributed to their endosymbiont. This would suggest that any selective pressure on the aphid would be on the endosymbiont as well. This in turn implies that any endosymbiont changes beneficial to its aphid host may result in more adaptive RWA biotypes.

The main aim of this study was to look at the role, if any, of the RWA endosymbiont, *Buchnera aphidicola*, on the RWA-wheat interaction. It was established that no secondary symbionts were present in the ten RWA biotypes investigated and that all the aphid biotypes only had their primary endosymbiont, *B. aphidicola*. Previous studies showed that this endosymbiont played a crucial role in aphids' ability to survive on plantsap by increasing essential amino acids that occurs in limited amounts within it. However, other reports on the RWA-*B. aphidicola* interaction suggested the endosymbiont played a decreasing role in RWA fitness in that it produced less essential amino acids. This hypothesis was further supported by studies that showed that the RWA had the ability to alter amino acid composition of susceptible wheat cultivars by increasing those essential amino acids previously supplied by its endosymbiont. However, in these studies, the RWA could not alter the amino acid composition of resistant cultivars. The results presented here suggest that *B. aphidicola* may still play a role in aphid fitness. Firstly, copy numbers of an essential amino acid producing plasmid varied amongst the different *B. aphidicola* originating from the different RWA biotypes. Secondly, though little sequence variation was found amongst these endosymbionts, the single CCC-insert difference did have a role in increasing the expression levels of the subsequent genes when compared to their plasmid copy numbers. Furthermore, the CCC-insert also increased the leader sequence length, thus resulting in more stable 5' UTRs. These findings, together with similar variations identified in this study within other aphid species, suggest a regulatory function for this region. However, they also suggest that *B. aphidicola* may play a role in RWA fitness.

Continuous pressure of low concentrations of essential amino acids in resistant wheat cultivars, selected for aphids that can survive under these limiting conditions. Normally, the RWA would increase these essential amino acids in its diet while feeding. Failing to do so would severely limit or even kill a RWA population. Another way of dealing with this problem is through natural selection of RWA individuals that can upgrade their diet, *i.e.* aphids that have *B. aphidicola* that

supply it with the required amino acids. This can be done by increasing the output of the essential amino acid biosynthesising pathways, either via increased copy numbers or through higher expression levels of these genes. Results in the present study showed that there were differences in an essential amino acid plasmid's copy numbers – to date it is believed to be the only way that the *B. aphidicola* can regulate its gene expression. However, a mutation preceding the leader sequence of a leucine plasmid gene led to differences in transcript levels. This suggests another method through which *B. aphidicola* could alter transcript levels and thus play a role in its host's fitness. Indeed, these changes in plasmid expression in the endosymbiont may potentially result in the development of new biotypes, especially since RWA biotypes are currently distinguished from each other based on the different resistant cultivars they can feed upon. Alternatively, *B. aphidicola* could at the very least allow the RWA populations to survive on resistant wheat cultivars until they've adapted to feeding on them.

The second part of the study looked at the influences of statistical normalization methods on the identification of differentially regulated probe sets, when RWA-plant interactions are investigated with Affymetrix GeneChip technology. The hypothesis of this section stated that there should be a consistent subset of regulated probe sets/geneIDs identified as differentially regulated, regardless of the initial normalization method employed, as long as the rest of the analyses were constant, and that this subset would be normalization-method-independent. Two Affymetrix RWA-wheat interaction datasets, that included the interactions of two South African and two US RWA biotypes with a total of six wheat accessions, were investigated using five different normalization and background correction methods (RMA, GCRMA, MAS5.0, PLM, VSN) at three different confidence levels (95, 99, 99.9 %), in the presence and absence of FDR/FWER correction.

The results illustrated that genes identified as differentially expressed were highly dependent on the specific normalization and background correction method employed. It was also shown that normalization and background correction method dependent biases could not be nullified by increases in confidence levels, but that this actually tends to increase these biases. Furthermore, FDR and FWER detection methods usually add to the normalization and background correction bias, except in normalization approaches that are very strict in their selection criteria. This supports a hypothesis that the genes identified as differentially regulated depend on the inherited statistical biases of the background correction and normalization methods employed, and that both FDR/FWER and increased confidence levels could actually further enhance this problem. A multiple normalization approach on the initial data is proposed to enable the identification of a probe

set subset that is normalization-method-biases-independent. These genes/probe sets should not be influenced by any of the different normalization method biases, and would therefore reflect the true biological differentially regulated genes.