

# DNA Microarray Experimental Design and Software Based Data Normalization and Analysis

Von der Naturwissenschaftlichen Fakultät der  
Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des Grades  
Doktor der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation  
von  
Marcel von der Haar (geb. Koch), M. Sc.

geboren am 26. Mai 1988 in Salzgitter Lebenstedt

2017

Referent: Prof. Dr. rer. nat. Thomas Scheper

Korreferent: Prof. Dr. rer. nat. Bernd Hitzmann

Tag der Prüfung: 04.09.2017

*Für Kathrin und Lara*

## Acknowledgements

I would like to thank everyone that contributed to this thesis through support and guidance. At first I wish to express my gratitude towards **Professor Dr. Thomas Scheper** for providing me not only with my research study but also for the confidence and goodwill invested in me in the time of this thesis but also in my previous student years. I would like to sincerely thank **Professor Dr. Bernd Hitzmann** for assuming the responsibility as co-lecturer as well as **PD Dr. Ulrich Krings** for assuming responsibility as examination chairman.

A heartfelt thank you to **Dr. Frank Stahl** and **Dr. Patrick Lindner** who sparked my interest in microarray technology and bioinformatics. Thank you for hours of guidance, so many interesting discussions, sincere interest in the subject and my progress as a scientist. Thank you also for supporting and enabling the publication of my research.

I want to thank **Martin Pähler** for his wisdom and patience when introducing me to practical microarray experimentation and the many answers to my many questions. **Friedbert Gellermann, Thorsten Stempel, Thorleif Hentrop, Ivo Havlik** and **Michael Dors** were a great support, providing me with new insights and solutions, alternative perspectives and interesting debates. Thank you for that. I would also like to thank **Martina Weiß** for her professional support regarding experimental and administrative issues. A special thank you to **Ulrike Dreschel** and **Cornelia Alic** without whom I would have surely gotten lost in the administrative jungle or sued for customs infringements.

Furthermore I wish to thank **Andre Jochums** for enlightening walks, **Daniel Marquard** and **Christian Lüder** for their patience and debugging prowess and **Tim Lücking, Christoph Busse, Jan König** and **Lukas Raddatz** for their support and company inside and outside the Braukeller. I wish to express my gratitude to the entire institute for their unequivocal and uncomplicated support and generosity at the time of the birth of my daughter **Lara**.

Most of all I wish to thank **my family** and my wonderful wife, companion and capable and inspiring colleague **Kathrin von der Haar**. Your never-ending support, warm-heartedness, understanding and confidence in me through all these years truly is the source of my accomplishments.

# Zusammenfassung

Viele der aktuellen Herausforderungen in Wissenschaft und Technik werden von zwei Themenkomplexen dominiert: Interdisziplinarität und Komplexität. In Medizin und Molekularbiologie stellt die Erfassung und das Verständnis der Gesamtheit der zellulären Prozesse eine solche Herausforderung dar. Ein Teilbereich dieses Themenkomplexes, die Transkriptomanalyse durch DNA Microarrays wurde in mehreren interdisziplinären Ansätzen in dieser Doktorarbeit analysiert und um neue Methoden erweitert.

Im ersten Teil dieser Arbeit wurde die photoneninduzierte Zerstörung weitverbreiteter Fluoreszenzmarker bei zweifarben-Microarrays charakterisiert. Das farbstoff- und scannerspezifische Photobleaching wurde quantifiziert und zur Erstellung eines empirischen Modells genutzt. Mit Hilfe dieses Modells lässt sich die durch Photobleaching eingetragene Verzerrung der Messdaten effizient minimieren.

Im zweiten Teil dieser Arbeit wurde einerseits eine technisch-chemische Schutzschicht für fluoreszenzmarkierte zweifarben DNA-Microarrays entwickelt. Sie basiert auf einem reduktiv-oxidativen Schutzpuffer und minimiert erfolgreich Photobleaching. Mit Hilfe dieses Schutzpuffers wurde andererseits ein weiterer verzerrender Effekt charakterisiert: Förster-Resonanz-Energie-transfer zwischen dem Donor Cy3 und dem Akzeptor Cy5. Dabei konnte nicht nur das allgemeine Auftreten des Effekts nachgewiesen werden, sondern auch seine Relevanz (Verzerrung) für die Analyse der resultierenden Scanning-Daten.

Der dritte Teil dieser Arbeit behandelt die Entwicklung einer .NET-basierten End-User-Software zur geführten Prozessierung, Normalisierung, Analyse und Veröffentlichung von DNA-Microarray Ergebnissen. Die Software enthält die neu erarbeiteten Normalisierungsmethoden aus den vorangegangenen Teilen dieser Arbeit sowie weitere Methoden. Neben der Normalisierung von Photobleaching wird das Vorhandensein desselben durch ein für low density Arrays trainiertes Neuronales Netz erkannt. Ein auf Varianzanalysen basierendes weiteres Modul evaluiert jede angewendete Normalisierung bezüglich ihres Effekts auf diverse Verzerrungen. Zusammen mit einem Tooltip-System wird dem Anwender so die Möglichkeit gegeben, unabhängig vom wissenschaftlichen Hintergrund fundierte Entscheidungen bzgl. der Behandlung seines Datensatzes zu treffen.

Alle Arbeitsteile dienen der Erhöhung von Vergleich- und Reproduzierbarkeit von DNA-Microarray Experimenten um deren Anwendbarkeit zu verbessern und das Potential dieser Technology auszuschöpfen.

**Schlagwörter:** Microarrays, Cyanin-Farbstoff, Photobleaching, FRET, ROXS, Analysesoftware

# Abstract

Many recent challenges in science and engineering are dominated by two subjects: Interdisciplinarity and complexity. In medicine and molecular biology, one of these challenges is the understanding of the entirety of cellular processes. In this thesis, a sub-area of this subject, DNA-microarray-based transcriptome analysis, was assessed and methodically advanced using multiple interdisciplinary approaches.

As a first of this thesis, photo induced destruction of fluorophore labeling agents widely used for two-channel microarray experiments was characterized. Dye and scanner-specific photobleaching was quantified and used to generate an empirical model. It was shown that by utilizing this model, photobleaching induced bias can be successfully normalized.

In the second part, as a technical, chemical solution for photobleaching minimization, a protective layer directly applied onto the array slide was designed. The protective function is based on a buffer including a reductive-oxidative system (ROXS). Employing this protective layer, another biasing effect could be characterized: Förster-Resonance-Energy-Transfer between the donor cyanine-3 and the acceptor cyanine-5. Next to verifying the existence of this effect in two-channel microarrays, its relevance of normalizing this effect prior to microarray data analysis was shown.

The third part of this thesis covers the development of a .NET-based end user software for guided processing, normalization, analysis and publishing of microarray experiment data. Previously mentioned novel normalization methods were implemented, together with other widely used methods. This software can not only normalize photobleaching, it also recognized it using a neural network trained with low density array data. A variance analysis based algorithm automatically evaluates each applied normalization regarding its effect on possible bias of various sources. In combination with a tool-tip system, it empowers the experimenter, regardless of scientific background, to make informed decisions regarding the individual handling of his microarray data set.

All parts of this thesis serve the improvement of comparability and reproducibility of DNA-microarray experiments to harness the full potential of this technology.

**Keywords:** microarrays, cyanine-dyes, photobleaching, FRET, ROXS, analysis software

# Table of Contents

1	Introduction.....	1
2	Theoretical Part.....	3
2.1	DNA Microarrays for Gene Expression Analysis.....	3
2.2	Cyanine Dye Labeling in Microarray Experiments .....	5
2.2.1	Photobleaching-Susceptibility.....	6
2.2.2	FRET - Förster-Resonance-Energy-Transfer .....	6
2.3	Principles of Microarray Data Processing and Analysis .....	8
2.3.1	The GPR Format.....	8
2.3.2	Filtering and Preprocessing.....	8
2.3.3	Normalization.....	9
2.3.4	Detection of Differential Expression .....	10
2.3.5	Software Solutions for DNA Microarray Experiment Analysis .....	13
3	Experimental Part.....	15
3.1	The Impact of Photobleaching on Microarray Analysis.....	15
3.2	Optimization of Cyanine Dye Stability and Analysis of FRET Interaction on DNA Microarrays .....	34
3.3	<i>Array Analysis Manager</i> – An automated DNA microarray analysis tool simplifying microarray data filtering, bias recognition, normalization and expression analysis .....	51
4	Conclusion.....	59
5	Literature.....	61
6	Figures .....	66
7	Tables .....	67
8	Licenses .....	68
9	Contributions.....	69
10	Author’s Resume.....	70

## List of Abbreviations

<b><math>\alpha</math></b>	Alpha
<b>ANN</b>	Artificial Neural Network
<b>ANOVA</b>	Analysis of Variance
<b>Cy3</b>	Cyanine-3
<b>Cy5</b>	Cyanine-5
<b>cDNA</b>	Complementary DNA (sometimes: copy DNA)
<b>DNA</b>	Deoxyribonucleic Acid
<b>dCTP</b>	Deoxy-Cytosin-Triphosphate
<b>dNTP</b>	Deoxy-Nucleotide-Triphosphate
<b>FDR</b>	False Discovery Rate
<b>FRET</b>	Förster-Resonance-Energy-Transfer
<b>GEO</b>	Gene Expression Omnibus
<b>GPR</b>	<i>GenePix</i> ® Results Format
<b>HSW</b>	Holm's Step-Wise Correction
<b>LOESS/LOWESS</b>	Locally Weighted Polynomial Regression
<b>MAQC</b>	MicroArray Quality Control Project
<b>mRNA</b>	Messenger RNA
<b>NCBI</b>	National Center for Biology Information
<b>NGS</b>	Next Generation Sequencing
<b>PBS</b>	Phosphate Buffered Saline
<b>PMT</b>	Photomultiplier Tube
<b>RNA</b>	Ribonucleic Acid
<b><math>R_0</math></b>	Förster distance
<b>ROXS</b>	Reductive-Oxidative System
<b>SAM</b>	Significance-Analysis of Microarrays
<b>SNP</b>	Single Nucleotide Polymorphism
<b>SNR</b>	Signal-to-Noise-Ratio
<b>SOP</b>	Standard of Practice
<b>SS</b>	Sum of squares
<b>ssDNA</b>	Single Strand DNA
<b>STR</b>	Short Tandem Repeat
<b>WPF</b>	Windows Presentation Foundation
<b>XAML</b>	Extended Active Markup Language



# 1 Introduction

*If we begin in certainties, we shall end in doubts; but if we begin with doubts, and are patient in them, we shall end in certainties.*

- Francis Bacon

*The manipulation of statistical formulas is no substitute for knowing what one is doing.*

- Hubert M. Blalock Jr.

In the 21<sup>st</sup> century, data has become a most valuable and abundant resource. Our ability to measure processes and generate data is the key stone to our advancement in all areas of science and engineering. The connection of data generation and the hope for advancement was almost never shown as clear as with the *Human Genome Project* in the 1990'ies. The characterization of all human genes, often referred to as "decryption", held out the prospect of great advancements e.g. in molecular biology and medicine. However, the sheer mass of generated data proved to be difficult to analyze. Instead of answering questions, the results generated many new ones. The idea of *genetic determinism*, a clear connection between genotype and phenotype, had to be mostly rejected. It was substituted with the realization that the phenotype is the result of a highly complex process of interactions between nucleic acids, proteins and elements of the cell plasma.

To illuminate these processes multiple new disciplines emerged, analyzing compositional changes of an organism biomolecules as a result of e.g. different stimuli. The term *Omics* was coined, summarizing observations of e.g. DNA (Genomics), RNA (Transcriptomics), proteins (Proteomics) and metabolites (Metabolomics). These disciplines depend on high-throughput technologies to allow for the parallel analysis of all genes, RNAs or proteins. In the case of transcriptomics, DNA microarray technology, next to NGS methods, is the method of choice.

DNA microarray experiments generate big datasets. Harnessing the information lying within this data necessitates sound knowledge of the underlying physical and chemical processes as well as proficiency in statistical paradigms. These interdisciplinary prerequisites in combination with the often varying scientific backgrounds of the experimenters lead to heterogeneous data generation and handling. This diminishes the transparency and reproducibility of these experiments and complicates the scientific discourse, limiting the applicability and capability of DNA microarrays.

This thesis aims to overcome these limitations with a dual approach. Firstly, generated microarray data quality is improved by optimizing the treatment and composition of the microarray slides,

## Introduction

---

the scanning setup and through the application of an empirical algorithm. Both strategies are designed to minimize/normalize bias introduced through bleaching of the labeling agents. This allows for the application of multiscan-techniques which in turn improves the dynamic intensity range the scanner can cover. Secondly a software was developed as part of this thesis. It is equipped with the aforementioned algorithms as well as an artificial neural network that allows for automatic photobleaching bias recognition and normalization. Additionally the software provides the user with ANOVA-based bias evaluations, so that he can make educated decisions on which normalizations should be applied to his dataset. These and other functions were developed to help harmonize the microarray data analysis process in order to expand the applicability of this technology.

## 2 Theoretical Part

### 2.1 DNA Microarrays for Gene Expression Analysis

The DNA microarray technology applies principles of nucleotide hybridization methods such as blotting and PCR as well as fluorescence-microscopy. It allows for high-throughput parallel analysis of large numbers of different nucleotide sequences. The main application is the gene expression analysis, where the regulation of multiple genes or even the whole genome is approximated through the evaluation of transcriptional changes. It is further used for genotyping of SNPs, STRs and more. One exemplary field of use is medical research and diagnostics. Here, the observation of transcriptional changes is used for clarification of disease processes or for diagnostic means [1,2]. Other fields include molecular biology or nutrition and food research [3].

The underlying principle of DNA microarray gene expression analysis is the competitive hybridization of differently dye-labeled cDNA-probes with spotted, immobilized DNA-targets (see Figure 2.1). Microarray slides are functionalized with spots of covalently bound nucleotide sequences. Each spot carries sequences designed to hybridize with cDNA derived from a single genes mRNA transcript. Prior to the hybridization mRNA acquired from different regulatory states of the chosen biological sample is transcribed to cDNA. Dye-labeling can take place co- or post-transcriptional and a different labeling agent is chosen for each regulatory state. Upon hybridization, cDNAs derived from the same gene transcript, but from different regulatory states compete passively and stochastically for the available suitable immobilized target sequences. Eventually, the intensity ratio of the dyes immobilized on a spot reflect the relative abundances of the respective gene's transcript in the different regulatory states under study. Depending on the number of genes that are evaluated in a microarray experiment, the arrays are referred to as "low density" arrays (up to several hundred genes) or "whole genome" array (all relevant genes of the organism under study). To allow for sufficient statistical power, DNA microarrays have multiple spots per gene (specified as "replicates"). Some setups work with so called "spike-in" spots. These spots hybridize special marker nucleotides of known quantities to allow for signal calibration [4,5].

Theoretical Part

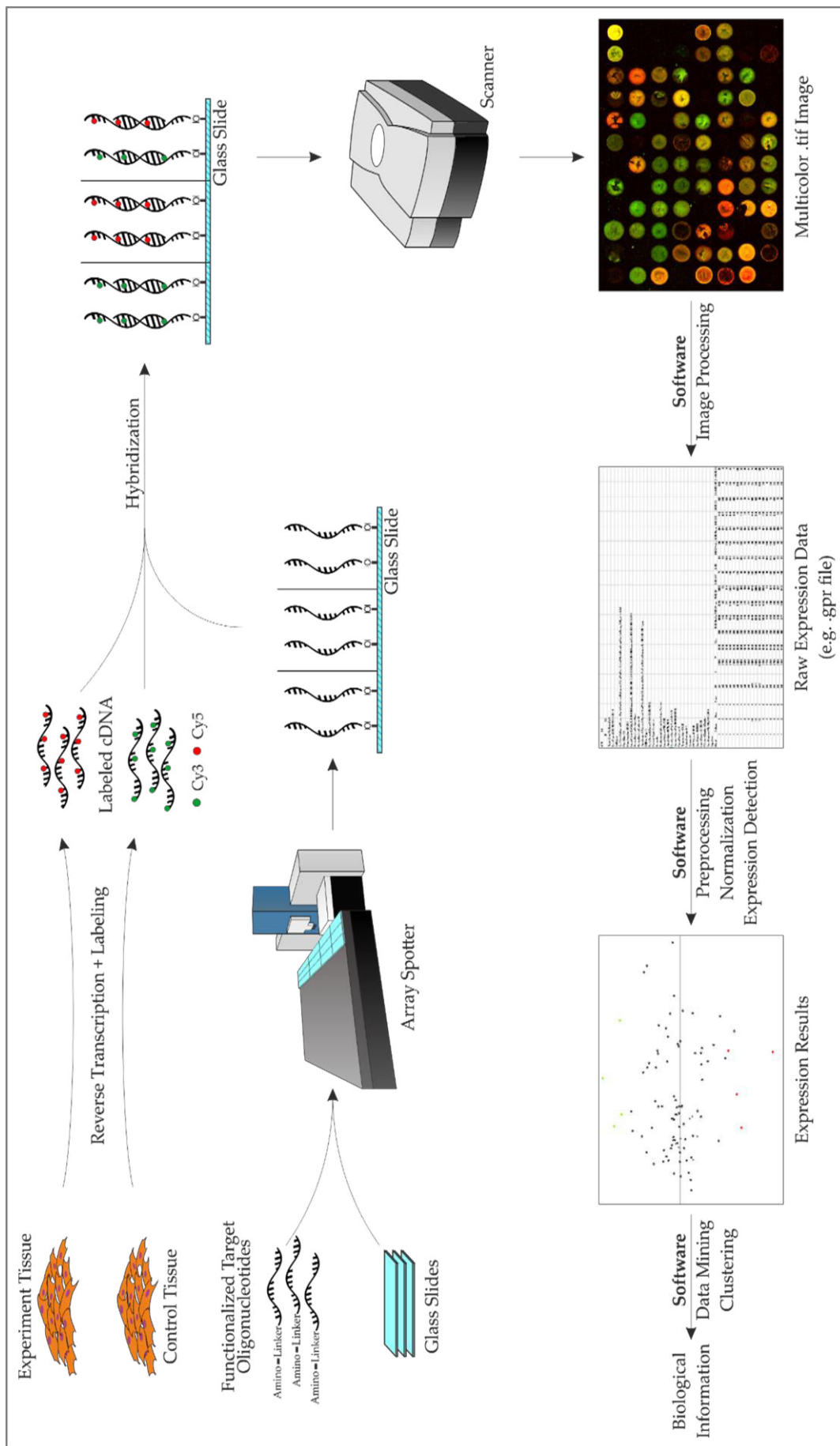


Figure 2.1 Schematic model of a typical DNA microarray experiment workflow.

## Theoretical Part

DNA microarray data is generated using laser scanners. The imaging is dominated by two processes (see Figure 2.2). Firstly, a laser beam irradiates the dyes, which are covalently bound to the immobilized oligos (see above). This induces the emission of lower energy photons from the dyes. The amount of irradiating primary photons is controlled by the applied scan power which can be set by the experimenter. The dye-emitted, secondary photons are not directly measured but their signal is transformed into an electronic signal by a photomultiplier tube. This electronic signal can be individually enhanced by setting the voltage applied to the photomultiplier's cathodes. As many DNA microarray experiments work with multiple dyes, settings such as the photomultiplier's voltages are set for each laser/dye individually in an attempt towards comparable dye signals [6].

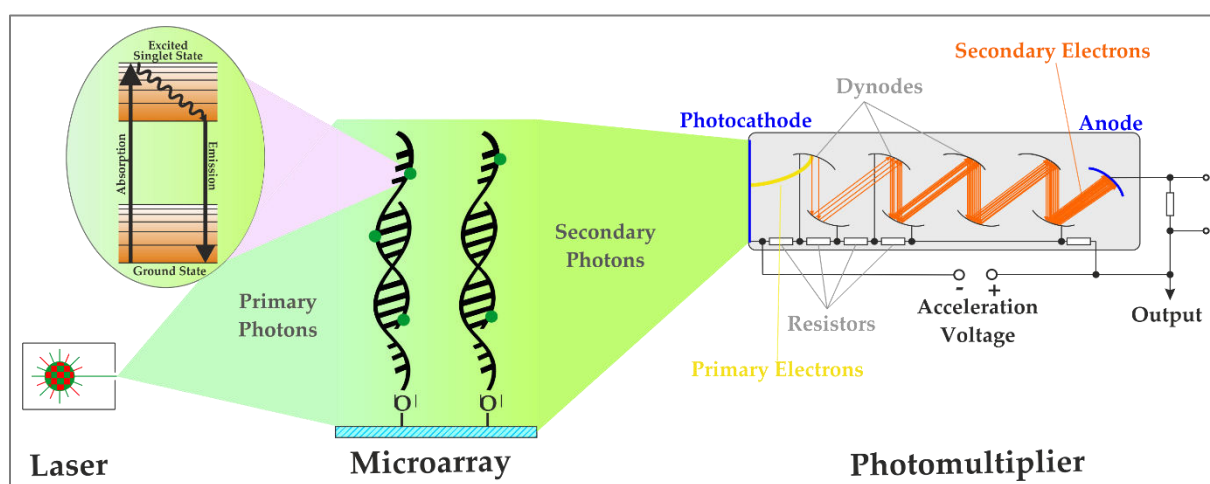


Figure 2.2: Schematic model of Cy3-fluorescence-labelling based DNA microarray scan imaging.

## 2.2 Cyanine Dye Labeling in Microarray Experiments

Today, experimenters can choose between many labeling agents [7]. While the number of dyes and labeling methods is ever growing, most approaches only allow for single channel labeling [8-11] or only allow for multiplexed qualitative labeling/detection [12-14]. The wish for comparable near-quantitative multi-channel labeling is one of the main arguments for the usage of cyanine fluorophore dyes. Especially Cy3 and Cy5 are a widely used combination [15]. These molecules only differ by one conjugated C-C double bond in structure and share many chemical and physical characteristics. Especially interesting for labeling uses is their small cross-talk [15,16] which should, in theory, allow for selective excitation of either Cy3 or Cy5 while their respective cyanine counterpart is present. Nonetheless, earlier works showed that the widespread application of

---

cyanine labeling can cause significant bias due to the dyes' differing susceptibility to photobleaching and through possible FRET interaction between Cy3 and Cy5 [17-21].

### 2.2.1 Photobleaching-Susceptibility

Photobleaching is an irreversible photochemical reaction which destructs a fluorophores ability to emit photons [22]. Photobleaching can be caused by photons and ozone [23]. The effect's magnitude is characteristic for each fluorophore [16,24,25]. A general schematic model of the underlying electronic processes can be seen in chapter 3.2, Figure 1. Many fluorophores are also excited by visible light, which is consequentially also a source of bleaching. This complicates the experimental handling in terms of reproducibility: extreme precision would be needed to guarantee that each DNA microarray is exposed to equal doses of visible light [25].

Several concepts were pursued in the past to minimize the photobleaching of cyanine dyes. Branham *et al.* proposed the installation of ozone filters in labs running microarray experiments and showed how ozone mediated bleaching can successfully be diminished by these efforts [26]. Some microarray scanner manufacturers already equip their scanners with ozone filters to minimize ozone bleaching in the scanning process [27]. A different approach tries to minimize the labeling agents' susceptibility towards bleaching by altering the molecule. Dar *et al.* modified Cy5, increasing its stability towards ozone and light to increase dye comparability in combination with the more photo-stable Cy3 [24]. Other takes on this subject tried optimizing scanner/laser settings and buffer composition [28]. Vogelsang *et al.* significantly reduced blinking and photo destruction of cyanine dyes by depopulating reactive intermediate states of the cyanine's excited electrons from which photo-destruction is initialized [20]. This was achieved by applying a buffer with oxidizing and reducing agents suited to catalyze the depopulation of critical states, based on works of Widengren *et al.* [29]. This increased dye longevity and quantum yields for experiments in aqueous solution using a fluorescence microscope [20,30].

The relevance of these effects for microarray experiments were shown by Satterfield *et al.* [25]. Monitoring intensity-changes of cyanine-3 (Cy3) and cyanine-5 (Cy5) serial dilution slides under heavy use over the course of five weeks, it was shown that multiple microarray scans also bleach the fluorophores. These findings are of even higher importance for experimental design that rely on multiple scans to extend the dynamic intensity range.

### 2.2.2 FRET - Förster-Resonance-Energy-Transfer

FRET (Förster/Fluorescence-Resonance-Energy-Transfer) described a physical energy transfer from an excited donor towards an acceptor [31]. This energy transfer distinguishes itself from most other transfers by being non-radiative. Instead, the energy is transferred through dipole-dipole coupling of the donor and acceptor molecules. This process is highly sensitive to changes

Theoretical Part

---

in donor-acceptor distance. The so called Förster-distance  $R_0$  is defined as the distance at which the transfer efficiency is 50 %. This sensitivity is also used as a means of determining molecular distances and structures [32]. Given the molecule densities that are usually found in DNA microarray spots [33], it is reasonable to assume, that cyanine labeled nucleotides are in a close enough proximity to allow for FRET to happen.

Previous studies showed that FRET occurs between the donor Cy3 and the acceptor Cy5 by observing “passive de-quenching”. This effects describes an observable increase in donor photon emission over time, caused by the increasing destruction of the acceptor, allowing a higher percentage of donor to emit photons instead of transferring their energy to an acceptor. “Passive de-quenching” of Cy3 by photo destruction of Cy5 was used to quantify FRET [34,35]. The relevance of a FRET bias for microarray experiments was demonstrated by Rao *et al.* [18]. In their research it was qualitatively evaluated if FRET is observable in DNA microarray two-dye experiments. This was done by partially exposing a spot containing Cy3- and Cy5-functionalized immobilized oligonucleotides to confocal laser light. Cy3 and Cy5 emissions before and after the exposure were compared and the expected anti-proportional change in intensity for both dyes was observed. Thus, the possibility of FRET biasing cyanine labeled microarray experiments cannot be discarded.

## 2.3 Principles of Microarray Data Processing and Analysis

### 2.3.1 The GPR Format

The microarray data on which this thesis is based was generated using the GenePix® scanner and software system. The GenePix® Pro software evaluates the scanned microarrays and converts the generated tiff-images of 2-channel fluorescence intensity into numeric data. This output data is provided in the GPR Format [36]. A GPR file consists of a header section containing general scanning information e.g. PMT voltages, laser and scan power, temperature and applied laser wavelengths. The results section is made up of a table containing various information for each spot previously defined and allocated by the experimenter. Available information includes foreground spot intensity means and medians (named feature intensities, integer values ranging from 1 to 65.535) as well as local background intensity means and medians for all used laser wavelengths. Measures of variability such as standard deviations, percentage of spot pixels outside of various thresholds and more are given. This information is complemented by quality indicators such as percentages of saturated pixels, percentages of pixel that are possibly affected by noise, the overall SNR and more. Additionally, GenePix® applies its own quality check algorithm and “flags” spots that fail their test. The validity of this algorithm has been positively evaluated by Repenning [37].

### 2.3.2 Filtering and Preprocessing

While commercial solutions such as the GenePix® platform apply their own filtering algorithm, and mark low quality spots e.g. with flags, the scientific community also developed multiple filtering approaches. For example, Lyng *et al.* recommend the exclusion of all spots showing median foreground intensities above 50,000 and below 1,000 to account for possible saturation and/or noise bias [38].

The log-transformation is a fundamental preprocessing step for microarray data. Next to the improved interpretability regarding biological issues, the transformation eliminates misleading disproportions between two relative changes [39]. In order to improve the statistical power and tackle noise issues, microarray experiments should be carried out with replicates. All spots and replicates of the same gene/feature can then be tested for outliers and finally be combined generating measure of central tendency (e.g. mean or median) and measures of variance. Typical outlier test are e.g. the Nalimov-Test, which is implemented in the *Array Analysis Manager* (see chapter 3.3).



### 2.3.3 Normalization

Basic normalization goals specific to DNA microarrays include the normalization of background and dye/color bias. Additionally, if an experimental design included multiple arrays, a between array normalization can be applied.

Several approaches to correct for background bias exist. GPR files provide the experimenter with local background intensity measures allowing for local background correction. Spot foreground intensity measure are corrected by simply subtracting the respective background. The sub-grid background correction calculates a measure of central tendency for all pixels in a sub-grid which is then used to allow for comparison between sub-grids. This method was developed with regard to spotting robots with pin-arrays, which produce iterative patterns. The group background correction uses all local background values of spots within a sub-grid to generate a measure of central tendency. In addition to these methods, experimenters use blank spots with no DNA for background correction or control spots with random nucleotide sequences to account for background and cross-hybridization bias [39].

As mentioned in chapter 2.2, the fluorescent dyes Cy3 and Cy5 are a preferred combination for two-channel experiments. In spite of their many advantages, these molecule are by no means perfect labelling agents. The size difference between these two could negatively affect the transcription rate of Cy5-labeled nucleotides in comparison to nucleotides labeled with the smaller Cy3. Furthermore, Cy3 and Cy5 show different correlations regarding the concentrations of hybridized labeled nucleotides vs. measured intensity as well as differential correlations between PMT voltage changes and intensity changes [6]. This differences can result in systematic distortions which are visible in Cy3/Cy5 scatterplots of whole genome data. The characteristic shape of this dye bias is the basis for normalization strategies [2,40]. This already leads to the main disadvantage inherent to these strategies. Relying on a visible distortion limits these approaches to datasets that actually display the bias. This is why these normalization methods should only be applied to whole genome array data. Method include curve fitting and piece-wise linear normalization techniques as well as linear or non-linear regression models such as LOESS and LOWESS, as used in the *Array Analysis Manager* (see chapter 3.3) [41]. Another way to account for dye bias, which can be applied to low density arrays as well, is by adjusting the experimental design. When comparing two regulatory states, the experimenter manufactures two arrays instead of one. The second array is hybridized with transcribed cDNA that was labeled complementary to the first array. This “dye swap” generates information of both used dyes for both regulatory states under study, minimizing dye dependent bias as both observations are biased comparably [42].

Array normalization methods are applied to normalize between array bias, allowing for comparisons of datasets derived from multiple arrays. Basic methods include dividing each gene's intensity information by the array mean of the same value. In the log-transformed context the array mean is obviously subtracted from each value. Similar to background normalization, approaches using control spots/genes can be used to compensate for between array bias. Analogous to the methods used for color normalization (iterative) linear regression methods also align values from different arrays [39].

Apart from these basic normalization effects various advanced methods have been suggested. For example, Bengtsson *et al.* developed an approach to normalize scanner specific bias for each laser channel using an extended dynamical range [43]. The use of an extended dynamic range itself is a method to normalize for bias introduced by saturation and/or noise effects. It is realized with multi-scan techniques. A microarray is scanned multiple times with varying PMT and/or scan power settings, the resulting datasets for each gene are filtered for a linear range of intensity vs. scanner setting correlations. This linear range is then used to calculate normalized intensities for each gene. This course of action is often rewarding as the dynamical range of actual spot intensities is significantly wider than the dynamic range a single microarray scan can cover. Without the use of multi-scan techniques information will be lost due to saturation and noise [43-45]. Tackling the same issue, Gupta *et al.* proposed a Bayesian hierarchical model to correct signal saturation and Yang *et al.* proposed a method for intensity estimation of spots with saturated pixels [46,47].

While many strategies have been proposed to normalize a vast array of bias from various sources, a harmonized SOP for microarray gene expression analysis that is widely used has not been established as of now. The MAQC I and II were first steps to gather information on the present state of microarray data generation and handling [48]. Allison *et al.* reviewed prevalent developments in microarray analysis, concluding that consensus is starting to emerge but still far from being established [2]. The resulting lack in comparability and reproducibility is a major hindrance for the application of DNA microarray experiments especially in clinical and commercial areas [49,50]. Whether a static SOP can be defined especially for experimental designs that do not use commercial microarray platforms is questionable.

#### 2.3.4 Detection of Differential Expression

Gene expression experiments are comparative studies. Most often 2 different states are compared, one state used as a reference or control. For instance, comparing healthy tissue with tissue affected by a disease e.g. cancer. Another typical example would be an organism under physiological conditions versus the same organism exposed to some form of stress (heat, toxins,

etc.). In DNA microarray gene expression experiments this comparison is quantified most often by the logarithmic ratio of foreground intensities. These foreground intensities may have been preprocessed and/or normalized (see chapters 2.3.2 and 2.3.3). This ratio is usually defined as follows:

$$\text{log-ratio} = \log_2 \left( I_1 / I_2 \right) \quad (1)$$

$I_1$ : mean/median foreground intensity of state 1

$I_2$ : mean/median foreground intensity of state 2

Once this value is determined for all observed genes, a method needs to be applied to separate genes depending on the absence or presence of expressional change. Optimally, a selection method should offer high sensitivity (few false negative decisions) and a high specificity (few false positive decisions) [39]. The simplest approach towards selection of differentially expressed genes is the fold change method. For this method a 2-fold or 4-fold threshold is chosen and genes are grouped in comparison of this fold change threshold. The inadequacy of this method is obvious, as no argumentation is provided on why a threshold of 2 or 4 has any biological or statistical significance. The experimenter has next to no control over sensitivity or selectivity. This method also tends to overestimate gene expression changes in lower more noise biased expression ranges while expressional changes are judged upon more conservatively in high expression ranges. A closer examination of the fold change method showed that the significance of fold change predictions is severely influenced by the threshold choice, which is problematic as no scientific or statistical reasoning is used to determine that threshold [51].

Another widely used method uses the mean experimental ratio and its standard deviation to select differentially expressed genes. The so called “unusual ratio” method determines the distance of each genes ratio from the experimental mean and set a threshold depending on the means standard deviation. Usually genes whose ratios differ by more than two mean standard deviations from the experimental mean are considered differentially expressed. While this method is still comparatively simple, it is superior to the fold change method, as the experimenter can apply statistical reasoning and set a threshold according to the significance level of his choice. This method is however limited as it relies on values derived from within the experiment. The significance of the experimental mean and of its variance decreases the smaller the number of observed genes gets. Another grave disadvantage lies within the statistical setup and is also more likely to affect low density arrays but could also affect whole genome arrays. Independent of the true number of differentially expressed genes, this method will always select the most changed genes in the dataset. Also it will always select a fixed percentage of genes depending on the chosen

## Theoretical Part

significance level. While for whole genome experiments, the experimenter can at least estimate a percentage of genes affected based on empirical knowledge, such an estimation cannot be made with clear conscience for low density arrays. Low density arrays only carry subsets of the genome that is affected by genetic regulation. This subset does not necessarily have to show patterns of expressional change analogue to the whole genome. The other way around, without further proof it is inadmissible to project whole genome characteristics onto subsets of this genome.

A better statistically backed approach to select differentially expressed genes uses univariate statistical tests. This approach has several advantages over fold change or unusual ratio methods. The experimenter can select a threshold  $\alpha$  to adjust the significance of the test according to the prerequisites of his experiment. Instead of comparing a genes log ratio with possibly biased experimental means and variance, hypothesis test based method compare the intensities that make up an individual genes log ratio. The advantage of being able to set a static  $\alpha$  is also a disadvantage, though one, that can be accounted for. As hypothesis test methods carry out tests for each gene individually, large numbers of tests are carried out, especially for whole genome arrays. Similar to applying the unusual ratio method, setting a static  $\alpha$  for all carried out tests will result in the selection of a certain percentage of genes as differentially expressed. In contrast to the usage of the unusual ratio, the individual testing does allow for the implementation of correction algorithms for multiple comparisons. The most basic approaches, such as the Šidák or Bonferroni correction, simply minimize  $\alpha$  depending on the number of tests carried out to obtain a new static value used for each test (see (2) and (3)).

$$\alpha_c = 1 - \sqrt[R]{1 - \alpha} \quad (2)$$

$\alpha_c$ : Šidák-corrected  $\alpha$

R: Number of tested genes

$$\alpha_c = \frac{\alpha}{R} \quad (3)$$

$\alpha_c$ : Bonferroni-corrected  $\alpha$

R: Number of tested genes

These approaches lead to highly conservative decisions, in other words many false negatives. Dynamic methods such as the FDR and HSW corrections, both implemented in the *Array Analysis Manager* (see chapter 3.3), provide higher sensitivity without a complementary loss in selectivity. These method apply dynamically adjusted  $\alpha$ s for each individual gene tested (see (4) and (5)).

$$\alpha_c(k) = \frac{\alpha}{R - k + 1} \quad (4)$$

$\alpha_c(k)$ : HSW-corrected  $\alpha$  of the k-th gene

R: Number of tested genes

k: Index of tested gene, ordered  
by increasing p-value

$$\alpha_c(k) = \frac{k}{R} \times \alpha \quad (5)$$

$\alpha_c(k)$ : FDR-corrected  $\alpha$  of the k-th gene

R: Number of tested genes

k: Index of tested gene, ordered  
by increasing p-values

As a first step, all genes are sorted based on their p-values. Depending on this sorting each gene is tested with an adjusted  $\alpha$ . Tests are carried out ascending or descending p-value order. Different abort criteria determine which genes can be considered differentially expressed or not [39]. Permutation-based methods such as the Westfall and Young step-down correction or SAM should only be applied if sufficient numbers of permutations are available, a prerequisite that is often unfulfilled, especially in human medical studies [52,53].

The selection of available methods is constantly growing [54]. Kerr et al., Ambroise *et al.* and others proposed ANOVA based univariate approaches [44,55,56]. Khan *et al.* developed sequential algorithms to validate microarray data analysis, while Clark *et al.* utilized multivariate statistics to determine differentially expressed genes [57,58].

### 2.3.5 Software Solutions for DNA Microarray Experiment Analysis

Many tools have already been developed allowing for automated handling of image analysis, clustering, normalization, pathway analysis, database management, data visualization and more. Most commonly, microarray data is processed and analyzed using tools such as *Partek* and *Spotfire* in addition to tools created by microarray platform manufacturers such as *GeneSpring GX*

(Agilent Technologies, Santa Clara, CA, USA) or *Genome Studio* (Illumina Inc., San Diego, CA, USA) [59,60]. Alternatively, experimenters rely on the use of more flexible and universal programs or even programming languages such as *R*, *Bioconductor* or *Excel* [61,62]. The former allow for a guided analysis process with appropriate but limited tools for a limited set of experimental designs. The latter enable the user in principle to analyze any dataset with a wide variety of tools but without any guidance on which process is appropriate. Especially commercial solutions designed by microarray technology providers such as *GeneSpring GX* are designed to work with whole genome arrays provided by the same manufacturer. Thus, available pre-processing and normalization methods are suited for their standardized whole genome arrays. Individual, low density array setups as often used in academic research environments are an unattractive target from a commercial stand-point. Additionally, many assumptions made to justify the application of statistical method and normalizations can only be made for whole genome arrays (see chapter 2.3.3). As a result, experimenters working with individual low density designs have a much more limited list of tools and methods to choose from. Instead of utilizing analysis suites, low density array analysis is most often based on individual solutions employing *Bioconductor*, *R* and others. As low density arrays lack whole genome like dataset characteristics, advanced methodology, based on multivariate analysis and machine learning is applied to successfully differentiate expressional changes [54,63-65].

### 3 Experimental Part

#### 3.1 The Impact of Photobleaching on Microarray Analysis

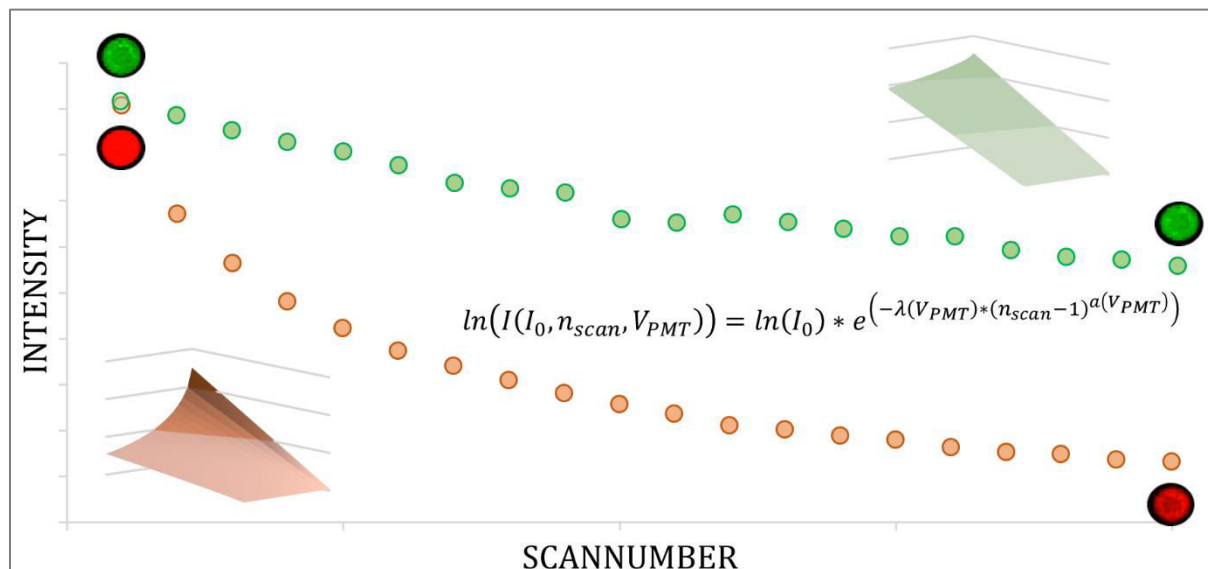


Figure 3.1: Graphical abstract of „The Impact of Photobleaching on Microarray Analysis“

The following section investigates the susceptibility of the fluorophores Cy3 and Cy5 towards photobleaching occurring in multiscan DNA microarray experiments. Intensity decreases resulting from bleaching are characterized for each dye individually. Findings of other work groups were validated, showing that Cy5 is significantly more affected by bleaching than Cy3 [24,25]. This differential behavior can result in significant bias, especially for multi-scan designs. In order to correct for this bleaching bias, an empirical model was devised. Based on the data generated in bleaching experiments, this model predicts the measured intensity loss of a spot depending on the applied PMT voltage, the initial spot intensity of the first scan, the choice of cyanine dye and the total number of performed scans. This model can be adjusted to infer the theoretical initial spot intensity for any spot, if information about number of previous scans is given.

Most multi-scan designs rely on the existence of a linear correlation between applied scanner settings such as the PMT voltage and the measured spot intensity [38,43,44,66]. These correlations are then used to compare spot information over several scans using e.g. a defined intercept. However, this course of action is legitimate only if all linear correlations determined show comparable slopes. In other words, the resulting straight lines should show a high degree of parallelism. The characterization of photobleaching in this study revealed that the reduction of measured intensity is in no way a linear process but can be modelled using a twofold exponential term. The bias introduced into the data by this process should therefore negatively affect the

Experimental Part

---

supposed parallelism of linear fits, especially as the degree of bleaching is depending on the initial intensity of a spot. In order to confirm that the devised correction model is actually improving the suitability of previously photobleaching-biased data for multi-scan purposes, its effects on the slope variance of linear fits was observed. The results of this inquiry clearly showed that the correction model significantly improves the data quality. The overall results of this study emphasize that photobleaching derived bias has to be accounted for, especially in multi-scan experiments. It also provides a mathematical solution that can be applied to microarray data as is.



Article

## The Impact of Photobleaching on Microarray Analysis

Marcel von der Haar \*, John-Alexander Preuß, Kathrin von der Haar, Patrick Lindner, Thomas Scheper and Frank Stahl

Institute of Technical Chemistry, Leibniz University Hanover, Callinstr. 5, 30167 Hanover, Germany; E-Mails: johnalexanderpreuss@googlemail.com (J.-A.P.); vonderhaar@iftc.uni-hannover.de (K.H.); lindner@iftc.uni-hannover.de (P.L.); scheper@iftc.uni-hannover.de (T.S.); stahl@iftc.uni-hannover.de (F.S.)

\* Author to whom correspondence should be addressed; E-Mail: koch@iftc.uni-hannover.de; Tel.: +49-511-762-2316; Fax: +49-511-762-3004.

Academic Editor: Chris O’Callaghan

Received: 29 June 2015 / Accepted: 8 September 2015 / Published: 11 September 2015

---

**Abstract:** DNA-Microarrays have become a potent technology for high-throughput analysis of genetic regulation. However, the wide dynamic range of signal intensities of fluorophore-based microarrays exceeds the dynamic range of a single array scan by far, thus limiting the key benefit of microarray technology: parallelization. The implementation of multi-scan techniques represents a promising approach to overcome these limitations. These techniques are, in turn, limited by the fluorophores’ susceptibility to photobleaching when exposed to the scanner’s laser light. In this paper the photobleaching characteristics of cyanine-3 and cyanine-5 as part of solid state DNA microarrays are studied. The effects of initial fluorophore intensity as well as laser scanner dependent variables such as the photomultiplier tube’s voltage on bleaching and imaging are investigated. The resulting data is used to develop a model capable of simulating the expected degree of signal intensity reduction caused by photobleaching for each fluorophore individually, allowing for the removal of photobleaching-induced, systematic bias in multi-scan procedures. Single-scan applications also benefit as they rely on pre-scans to determine the optimal scanner settings. These findings constitute a step towards standardization of microarray experiments and analysis and may help to increase the lab-to-lab comparability of microarray experiment results.

**Keywords:** microarray; DNA; photobleaching; fluorophore; cyanine dye; bioinformatics; bioanalytics

---

## 1. Introduction

DNA microarrays have become a powerful tool for systematic monitoring of gene regulation. The technology is based on the competitive hybridization of differentially fluorophore-labeled cDNA-probes with spotted, immobilized DNA-targets. The cDNA's are transcribed from mRNA acquired from different regulatory states of the chosen biological sample. Thus, the ratio of the immobilized fluorophores on a spot reflects the relative abundance of RNA of the regulatory states under study. Within the last two decades the aforementioned principle has gained widespread use in fields such as molecular biology, genetics, and medicine [1,2]. It allows for the high-throughput transcriptome analysis of transcriptome regulation from a few dozens of genes up to the whole genome of the organism of interest [3].

The vast possibilities this technology provides are evenly met by technical, biochemical, and statistical difficulties. Each step of a microarray experiment introduces new factors that influence and possibly bias the final data. Beginning with choice of sample recovery and primer design, which might cause sequence-dependent bias [4]. Furthermore, the used spotting technique, as well as the choice of buffer, spotting, incubation and washing conditions, all influence spot geometry and uniformity by affecting drop drying and hybridization efficiency [5–9]. Data acquisition is facilitated using laser scanners controlled by PC software. Here, influencing factors are the scanner and its lasers themselves [10–12], the choice of fluorescent dye [13] as well as the scan settings, especially the scan power and the photomultiplier tube's (PMT) voltage [14,15], and also exposure to environmental light, ozone, and laser light prior to the data acquisition [11,16,17]. While this multitude of factors does not hinder the acquisition of significant data, it is a major barrier for lab-to-lab reproducibility, comparability, and consistency of microarray experiment data [18].

In order to overcome these limitations a vast array of tools has been developed. Some factors are addressed by changing the experimental design, e.g., additionally using reverse dye assignments (*dye swap*) to account for dye bias [19]. Several techniques focus on the data acquisition itself. Finding the optimal scanner settings has been the subject of a lively discussion [14]. Regardless of the respective settings, all single scan approaches suffer from a limited dynamic range of measured intensity, as the dynamic range of fluorescence intensity exceeds the dynamic range of a single array scan by far [14]. Two basic approaches have been suggested to overcome these limitations. Mathematical or statistical approaches try to correct for saturation or noise using information inherent in the acquired data. Gupta *et al.* [20] for example devised a Bayesian hierarchical model that corrects signal saturation based on pixel intensities. Most approaches however extend the scanning routine by recording multiple scans with different settings. The benefits of multiscan techniques for extending the linear signal range were, among others [21], shown by Khondoker *et al.* [10] who are using a maximum-likelihood-estimations model based on a Cauchy distribution to account for saturated signals and systematic bias. Ambroise *et al.* [12] characterized a PMT independent optical scanner bias that takes account for scanner specific bias. Based on this, a two-way ANOVA model was devised that accounts for scanner bias as well as saturation and noise through utilization of

multi-scan data. Multiscan techniques were shown to increase overall data quality as well as reproducibility in comparison with single scans [15,22]. They can also be used to normalize dye specific bias as an alternative to limited methods based on LOESS/LOWESS and others [14,21,23].

An ubiquitous difficulty when working with fluorophores is photobleaching, an irreversible photochemical reaction which destructs the fluorophores ability to emit photons [24]. Photobleaching is caused by photons and ozone and differs from fluorophore to fluorophore [11,16,25]. Satterfield *et al.* [11] showed that microarray scans also bleach the fluorophores when they monitored intensity-changes of cyanine-3 (Cy3) and cyanine-5 (Cy5) serial dilution slides under heavy use over the course of five weeks. These findings imply a possible effect of photobleaching on multiscan data quality.

In this study, we evaluate the photobleaching characteristics of Cy3 and Cy5 as part of solid state DNA microarrays. The effects of initial foreground intensity on the degree of bleaching as well as the effect of laser scanner dependent variables such as the PMT voltage on the imaging are investigated. Several microarray slides with identical layout were manufactured with conditions optimized in a previous study and repeatedly scanned with individual static PMT voltages. Identical 5'-cyanine functionalized single strand DNA was immobilized onto the slides in order to reduce sources of bias, such as the sequence differences or dye-incorporation and hybridization efficiency. The resulting data is used to develop a mathematical model capable of predicting the expected degree of signal intensity reduction caused by photobleaching for each fluorophore individually, depending on the initial foreground intensity, the number of previous scans and the desired PMT voltage in order to allow for the removal of photobleaching-induced, systematic bias in multi-scan procedures.

## 2. Model

Microarray scan imaging is dominated by two processes. Firstly, the immobilized, dye functionalized oligos are irradiated by a laser beam, which induces the emission of lower energy photons from the dyes. As the applied scan power is not varied in this study no closer look is taken at the relation between applied power and dye-emitted photons. However, considering photobleaching, this process is of utmost interest, as the cyanine dye loss of photo activity is photon-induced. Although the mechanism is not completely understood yet, it can be assumed that bleaching affects each cyanine molecule independently. Also, not every excited molecule is bleached. This leads to the assumption that photobleaching can be described as a degradation process, analogue to radioactive decay:

$$p(p_0, n_{scan}) = p_0 \times e^{-\lambda \times (n_{scan}-1)} \quad (1)$$

where  $p(p_0, n_{scan})$ : photons emitted after  $n$  scans;  $p_0$ : initial photons emitted ( $n_{scan} = 1$ );  $n_{scan}$ : number of scans;  $\lambda$ : degradation constant (neglecting a change of scan power,  $\lambda$  is assumed to be dye specific).

The photons, emitted from the cyanine dyes, are not directly measured by an optoelectronic transducer. They pass the PMT, which acts as a signal enhancer and transducer. In this vacuum tube, the photons strike a photocathode and, as a consequence of the photoelectric effect, electrons are ejected. These electrons again strike a dynode that acts as a multiplier, emitting more secondary electrons. Several dynodes work as a cascade, each holding a higher positive potential than its predecessor and each multiplying its predecessor's electron signal. Finally, the secondary electrons strike the anode, where the signal is transduced. The extent of signal amplification depends on the voltage setting of the PMT. As

multiscan techniques are designed to enlarge the linear signal range of microarray experiments through variation of PMT voltages, it is crucial to characterize and model the PMT voltage's influence to fully understand its effect on imaging of photobleaching. As a consequence of the previously described cascade effect, the PMT signal enhancement is modeled by an exponential function, similar to Khondoker *et al.* [10]:

$$I_e(p_0, n_{scan}) = e^{\times p_0} \times e^{-\lambda \times (n_{scan}-1)} \quad (2)$$

where  $I_e(p_0, n_{scan})$ : post PMT intensity (electron signal);  $p_0$ : pre PMT intensity (photon signal, theoretical, not measured).

The above model involves a significant problem:  $p_0$ , the emitted photons of the first scan cannot be measured directly. The closest to  $p_0$  is  $I_0$ , the post PMT electron signal of the first scan. As described above, the electron signal is an exponential transformation of the photon signal. The exponential relationship cannot be exactly determined. However, transforming the relationship into a linear one by using the natural logarithm of  $I_0$  instead constitutes a practical solution. A model calculated with  $\ln(I_0)$  is valid as long as one stays in the  $\ln(I_0)$ -based reference system:

$$\ln(I(I_0, n_{scan})) = \ln(I_0) \times e^{(-\lambda \times (n_{scan}-1))} \quad (3)$$

where  $I(I_0, n_{scan})$ : post PMT foreground intensity after n scans, with given  $I_0$ ;  $I_0$ : initial post PMT foreground intensity ( $n_{scan} = 1$ );  $n_{scan}$ : number of scans;  $\lambda$ : degradation coefficient.

At this time, the model does not directly feature the applied PMT voltage. It might not have to directly incorporate the voltage at all if its influence is already sufficiently covered by  $I_0$ , which itself is directly dependent on the applied PMT voltage. In case that our model does not account for all major variance in the data an additional parameter is introduced. This parameter must be consistent with our degradation or decay model, e.g., the model should return  $I(n_{scan}) = I_0$  for  $n_{scan} = 1$ . This condition rules out intercepts and coefficients on the linear level of our model. The exponential term cannot be extended by adding an intercept for the same reason. The addition of an exponential coefficient would be redundant as one already exists ( $\lambda$ ). Adding an exponent to  $(n_{scan} - 1)$ , however, allows for the alteration of the degradation behavior without thwarting the conditions of a degradation model.

The combination of models (1), (2) and (3) together with the abovementioned considerations lead to the following function, which is theoretically suited to model the effect of photobleaching on measured intensities of microarray scans, taking into account the initial measured intensity ( $I_0$ ), the number of previously executed scans ( $n_{scan}$ ):

$$\ln(I(I_0, n_{scan})) = \ln(I_0) \times e^{(-\lambda \times (n_{scan}-1)^a)} \quad (4)$$

where  $I(I_0, n_{scan})$ : post PMT foreground intensity after n scans, with given  $I_0$ ;  $I_0$ : initial post PMT foreground intensity ( $n_{scan} = 1$ );  $n_{scan}$ : number of scans;  $\lambda$ : degradation coefficient;  $a$ : exponent.

### 3. Experimental Section

#### 3.1. Oligo Preparation

Single strand DNAs (ssDNA) of 40 nt length were purchased from Eurofins Genomics GmbH (Ebersberg, GERMANY). The internally-compiled sequence was optimized with regard to low stabilities of potential homodimers and hairpins. The 5'-end of the ssDNA was modified with a Cy3 or Cy5 respectively.

The 3'-end of the ssDNA was modified with an amino-modified C7 spacer: 5' Cy3/Cy5–C ACG ATT CGG CTT TAG GTC AAC TGG ATT TCG GCT TAG GAC–C7-Amino 3'. In order to minimize variance it was decided to use only one sequence, with one spacer-type and a set dye abundance per oligo. Instead of a real hybridization, both Cy5 and Cy3 dyes on nt-identical but mixed DNA pools are printed together as sequence-identical ss-DNA 40-nt strands. While this does not reflect the realities of an actual microarray DNA hybridization experiment, it is suitable to demonstrate the effect of photobleaching as well as it can be used as the basis for quantification. Each oligo was serially diluted with a buffer containing 3× standard saline citrate (SSC) and 0.001% 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS) to concentrations ranging from 5 to 0.05  $\mu\text{M}$  (a detailed table can be found in the supplementary materials). The buffer composition was chosen as a result of preliminary tests based on the works of Dawson *et al.* [6] in order to allow for homogenous distribution of the spotted oligos and minimized drying effects, thus minimizing spot heterogeneity (spot homogeneity information can be found in the supplementary materials). Solutions were stored at 4 °C and protected from light.

### 3.2. DNA Immobilization

DNA sequences were immobilized on the aldehyde glass slides (SuperAldehyde 2; Arrayit<sup>®</sup> Corporation, Sunnyvale, CA, USA) using a non-contact-spotter (Nano Plotter<sup>™</sup> NP2.1; GeSiM mbH, Großberkmannsdorf, GERMANY) with an applied voltage of 75 V. The selection of a contactless printer allowed for higher homogeneity in spot geometry by avoiding pin-derived variance. Concentrations between 0 and 5  $\mu\text{M}$  per dye were spotted in various pre-mixed combinations (a detailed table can be found in the supplementary materials). The spotting layout consisted of 2 × 8 blocks, where each block held 1 spot per oligo mixture giving a total of 16 spatially distributed spots per oligo mixture per slide. After drying the slides overnight in the dark, six washing steps using 4× SSPE buffer and water were performed, according to Dawson *et al.* [6].

### 3.3. Data Acquisition

All scans were performed using the GenePix<sup>®</sup> 4000B Microarray Scanner by Molecular Devices (Sunnyvale, CA, USA). All data was collected at a pixel size of 10  $\mu\text{m}$  and a total resolution of 1891 × 2089 pixels. Spot sizes were 229.48  $\mu\text{m} \pm 18.77 \mu\text{m}$ . Model data was acquired subsequently through one preliminary scan to determine the scan area and 20 additional scans per slide with constant PMT settings at 100% scan power, leaving approx. 6 min between the start of two scans. In this first modeling approach it was decided to only use 100% laser power in order to maximize the observable effect. Each slide was scanned with a different PMT setting, displayed in Table 1. Data collection was carried out by using GenePix<sup>®</sup>Pro 6.0 (Molecular Devices, Sunnyvale, CA, USA).

**Table 1.** PMT settings of different DNA chips.

# Chip	PMT <sub>635 nm</sub> [V]	PMT <sub>532 nm</sub> [V]
1	950	700
2	850	600
3	750	500
4	650	400
5	550	300

Validation data was acquired subsequently through one preliminary scan to determine the scan area and five additional scans with varying PMT voltage settings at 100% scan power (see Table 2). This independent data set consisted of three chips that were, except for the scanning process, identical in layout and processing to the five model chips.

**Table 2.** PMT settings of validation data.

# Scan	PMT <sub>635 nm</sub> [V]	PMT <sub>532 nm</sub> [V]
1	550	300
2	650	400
3	750	500
4	850	600
5	950	700

### 3.4. Data Analysis

#### 3.4.1. Post Processing

In addition to the criteria applied by GenePix<sup>®</sup>Pro in order to flag and exclude low quality spots, all spots with any saturated pixels as well as spot whose signal to noise ratio (SNR) was 3 or lower were excluded from further analysis. The SNR is defined as follows:

$$SNR = \frac{m_{Foreground} - m_{Background}}{s_{Background}} \quad (5)$$

where  $m$ : median;  $s$ : standard deviation.

Furthermore, following Lyng *et al.*'s recommendations [15], all sets of spots with median foreground intensities of the first scan ( $I_0$ ) above 50,000 and below 1000 relative intensity units were excluded from further analysis to prevent saturation and/or noise bias. Although a correction for background is a general convention, the actual application varies. Background correction is carried out locally, within a sub-grid, with blank spots or control spots. Most of these approaches have different underlying assumptions on how the background intensity reflects an intensity bias over- or better underlying the feature intensity. Furthermore Qin *et al.* [26] showed that while a background subtraction actually reduces the bias it increases data variability. Furthermore we have to investigate if and how the background intensity changes with increasing scans. If the background is indeed affected the question if the process occurs comparably on the surface of the actual spot still remains. These aspects were the basis of our decision to omit a background correction and to postpone a thorough examination of background photobleaching to future studies. Data conversion and filtering was carried out using the open source program R Studio Desktop v0.99.441 (R Studio, Boston, MA, USA).

#### 3.4.2. Modeling

The processed data was modeled using internally-written scripts in MATLAB v7.12.0.635 (The MathWorks, Inc., Natick, MA, USA).

This model concentrates on actual detected intensity and not on spotted concentration. This decision was made regarding intensity profile heterogeneity of replicate spots of the same concentration (e.g., for

Cy5 in this experiment, the average percent intensity deviation for replicate spots was approx.  $28.58\% \pm 20.17\%$ , more information can be found in the supplemental materials). This is a valid approach as the photobleaching depends on the actual amount of bound fluorophore on the spot and working with the intensity instead of the applied concentration allows for modeling without spot intensity profile bias.

At first, a regression was calculated for each independent spot, using the model described in Section 2 for both the Cy5 and Cy3 channel. For these regressions MATLAB's own non-linear least-square fitting algorithms based on trust regions was applied. Using Cy5 model data with  $R^2 \geq 0.95$  the dependency of both calculated parameters,  $\lambda$  and  $a$ , on PMT voltage and/or initial intensity was examined. Each variable, voltage and intensity, was examined independently for each parameter ( $\lambda$  and  $a$ ) by carrying out an analysis of variance (ANOVA). This approach was chosen to determine if a dependency can be observed that introduces a variance into the data, significantly higher ( $\alpha \leq 0.01$ ) than the experimental variance for the parameters ("Lack of Fit Test"). To allow for ANOVA analysis of  $I_0$  dependency,  $I_0$  data was organized in groups spanning 100 relative intensity units. Each significant dependency was then modeled using second order polynomials.

The acquired Cy5 model parameters were used to calculate a surface fit with the processed Cy3 model data. Cy3 parameters were modeled analogous to their Cy5 counterparts.

### 3.4.3. Validation

The generated models for both Cy5 and Cy3 photobleaching were applied onto the validation data, which was also processed as described in Chapter 3.3.1. The model term was converted to allow for the calculation of the initial intensity, given the current intensity ( $I(n_{scan})$ ), the used PMT voltage, and the amount of scans carried out before. The mean  $R^2$  of the linear fits of intensity vs. PMT voltage, as well as the standard deviations of the two linear parameters for all uncorrected data series were compared to the same criteria of all corrected data series for each cyanine dye independently.

## 4. Results and Discussion

### 4.1. Regression Analysis

Using model (4) with the preprocessed model data (exemplary shown in Figure 1) resulted in different outcomes for the two color channels. While for Cy3 59.9% of 1331 regressions had an  $R^2$  of 0.9 and above, 96.5% of all 1772 calculated regressions for Cy5 showed  $R^2$  of 0.9 and higher. This discrepancy could be a consequence of the well-known higher background of the Cy3 channel. The model data, however, contradicts this assumption as standard deviations for both channels are of comparable order and the SNR of the Cy3 channel is even higher (132.49) compared to the Cy5 SNR (51.11). Although Staal *et al.* [25] quantified the crosstalk of Cy5 to Cy3 as little as 0.2%, it is still possible, especially at higher PMT settings, that Cy5 crosstalk biases the Cy3 data. As the recorded spots were made of mixtures with varying concentrations of each dye, a spot with a high Cy5 concentration and a low concentration of Cy3 is likely to be biased in a more severely manner. An effect biasing the data could be Förster Resonance Energy Transfer (FRET) between Cy3 and Cy5 and intra spot heterogeneity. The transfer of energy between a donor and an acceptor in close proximity has been well described for nucleotide-bound fluorophores in general, and Cy3 and Cy5 specifically, [27,28]. Through FRET some

of the excited cyanines could have transferred the energy to their cyanine counterpart instead of emitting photons, thereby reducing the detected intensity of the respective channel. As FRET is highly dependent on a close proximity of donor and acceptor, this effect will be much more prevalent in high concentration spots or areas of higher nucleotide density in heterogeneous spots. The interdependency of FRET, intra-spot heterogeneity and photobleaching has been investigated by Rao *et al.* [29,30]. Radial and vertical intra-spot heterogeneity of printed targets profoundly influence local hybridization efficiency and finally the fluorescence signal as well as the occurrence of FRET. The described conjunction could also affect photobleaching rates as the excitation of one cyanine also partially excites the other one, thereby intertwining the exposition to potential photodestruction. Again the possible effect grows depending on the donor and acceptor concentrations. Furthermore Rao *et al.* [29] showed that the destruction of the FRET acceptor (here Cy5) leads to increased emission from the former donor (here Cy3), another source of signal crossover. The process of target-probe hybridization is the major influence modulating the scale of the phenomenon described before. This study's experimental setup relies on ssDNA printing of directly labeled nucleotides and no hybridization. While FRET and intra-spot heterogeneity can be expected to affect this data as well, the effect of hybridization cannot be accounted for and was subsequently not modeled. Although choice of experimental design regarding FRET complicates the generation of the Cy3 model, it shows that the usage of Cy3 and Cy5, although omnipresent in fluorophore-based bioanalytics, entails limitations that have not yet been properly addressed.

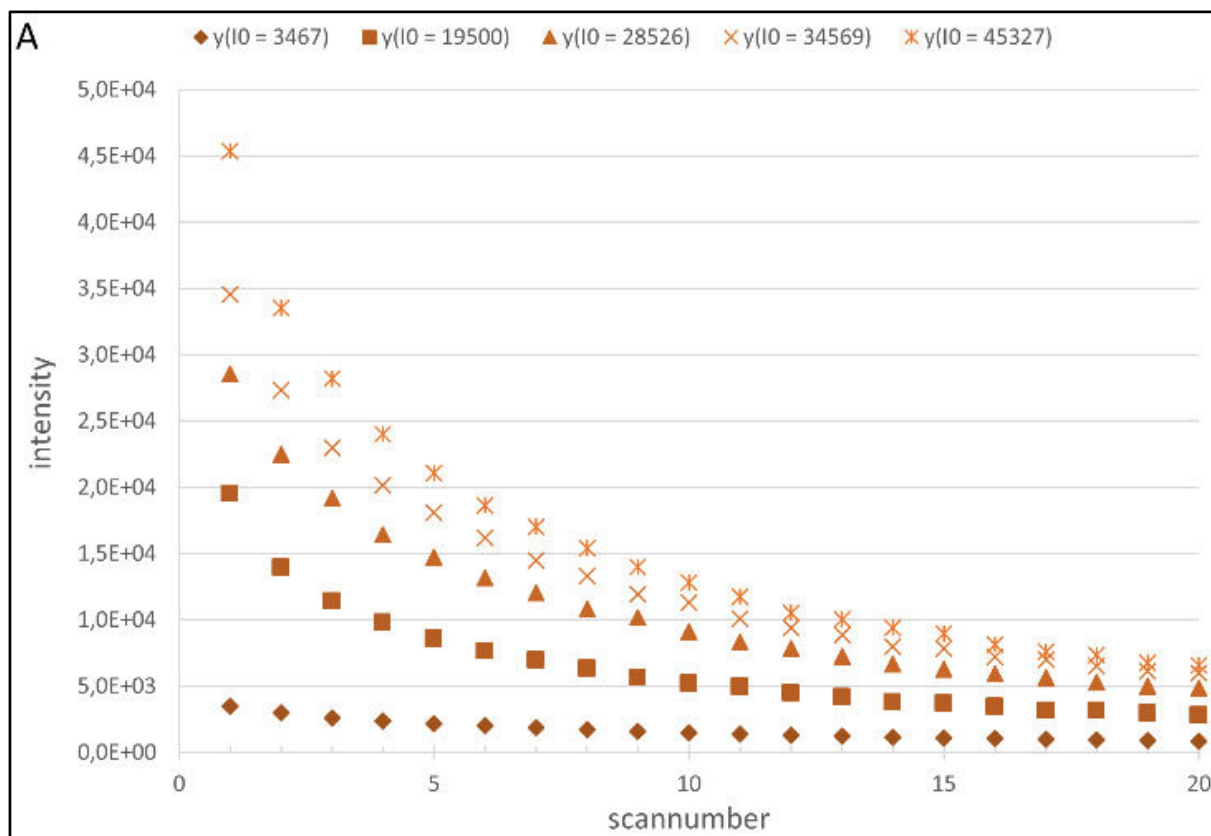
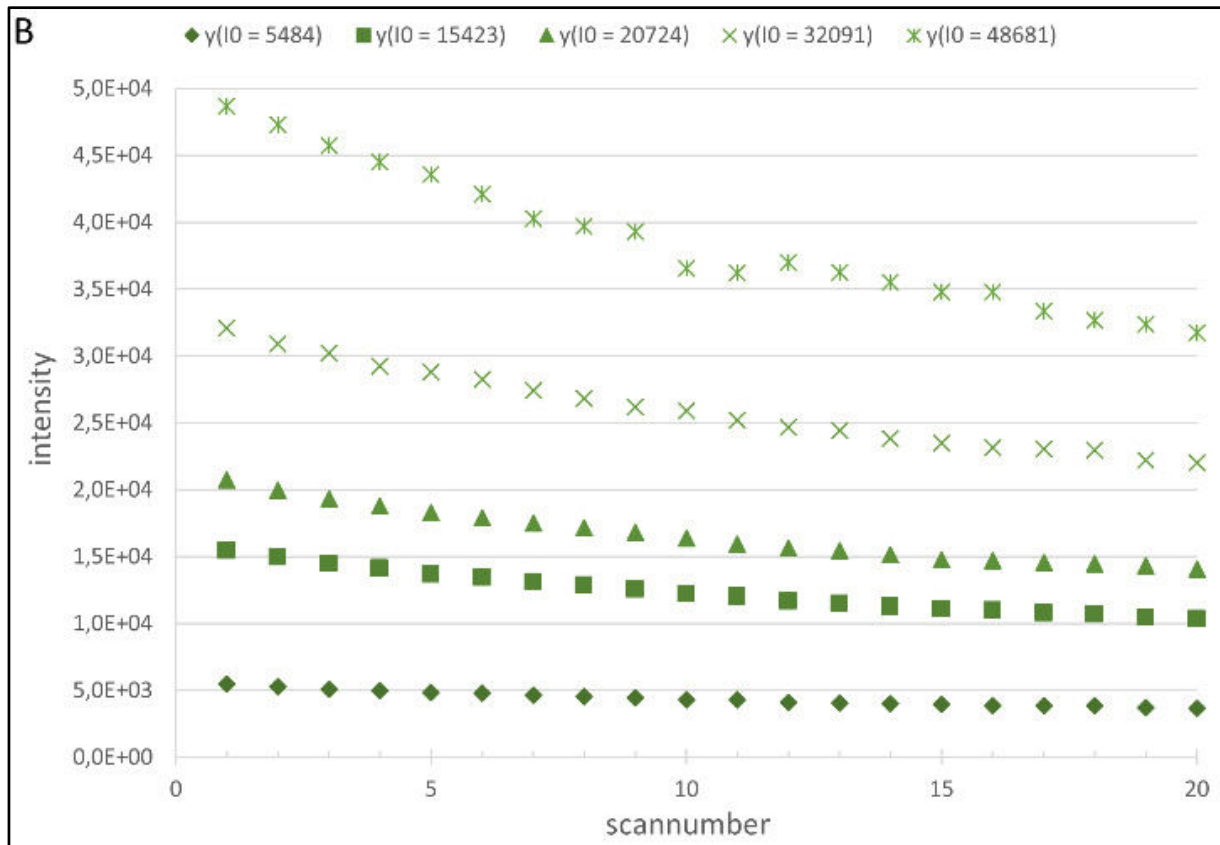


Figure 1. Cont.



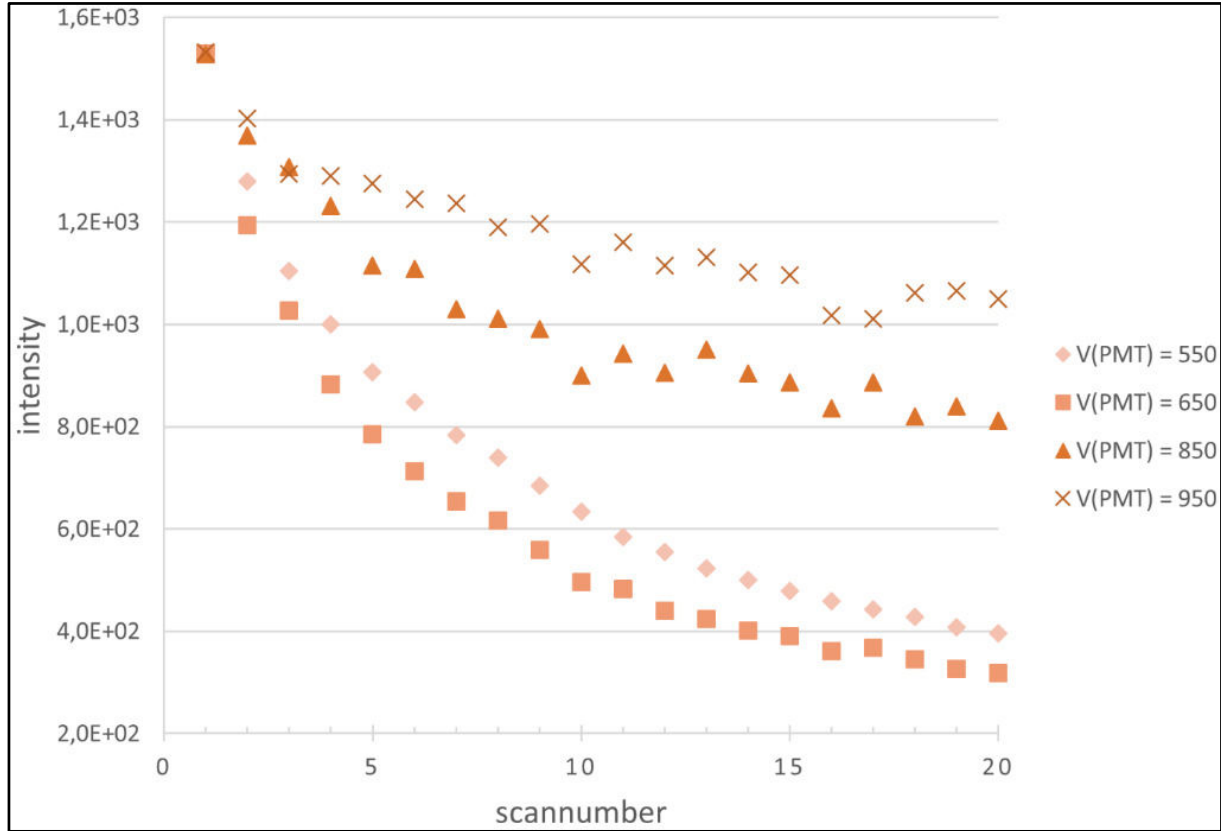


**Figure 1.** (A) Change in measured intensity of Cy5-labeled cDNA spots with increasing number of scans, depending on their initial intensity; and (B) change in measured intensity of Cy3-labeled cDNA spots with increasing number of scans, depending on their initial intensity.

#### 4.2. Generation of the Cy5 and Cy3 Model

With respect to these results, it was decided to focus on the Cy5 data for closer examination and to base a more refined model on this data. 96.5% of all 1772 calculated regressions for Cy5 showed  $R^2$  of 0.9 and higher and were used to generate the model. A model adjusted for Cy3 is calculated based on the Cy5 model. In order to investigate possible influences of the initial foreground intensity ( $I_0$ ) and/or the PMT voltage ( $V_{PMT}$ ) on both the degradation coefficient  $\lambda$  and the exponent  $a$ , multiple analyses of variance (ANOVA) were carried out. The underlying idea is to determine if the variance introduced to the parameters by the variables is significantly distinguishable from the experimental variance. This is a practical approach that does not ask if the variables actually influence our parameters, but if the modeling of any hypothetical influence can significantly improve the accuracy of the model, given the inherent experimental variance of the parameters. Firstly, the influence of  $I_0$  was investigated: Regarding  $\lambda$ , the null hypothesis ( $h_0: \sigma^2_{\text{model}} = \sigma^2_{\text{experiment}}$ ) cannot be rejected for any reasonable significance level  $\alpha$  ( $\alpha_{h_0 \text{ rejected, min}} = 0.9477$ ). For  $a$ , the lowest significance level that allows for rejection of  $h_0$  is even higher ( $\alpha_{h_0 \text{ = rejected, min}} = 0.9999$ ). As a result, both parameters are not modeled with regard of  $I_0$ . For  $V_{PMT}$ , however, results were different:  $h_0$  for  $\lambda$  as well as for  $a$  are rejected at an  $\alpha$  well below all levels established in applied statistics ( $\lambda: \alpha_{h_0 \text{ = not rejected, max}} = 9.09 \times 10^{-123}$ ,  $a: \alpha_{h_0 \text{ = not rejected, max}} = 4.08 \times 10^{-112}$ ). It is contradictory that  $V_{PMT}$ , a variable of a process succeeding the actual bleaching, is supposed to influence the parameter characterizing it. We assume that the PMT voltage's influence on  $\lambda$  does obviously not

display its influence on bleaching itself. A  $V_{PMT}$ -dependent  $\lambda$  is an expression of the transformation of the “observed” bleaching through the imaging process, which itself is  $V_{PMT}$ -dependent. These findings indicate that the variance introduced to the model data through  $V_{PMT}$  cannot be completely modeled indirectly using  $I_0$  alone, which is directly  $V_{PMT}$ -dependent. The effect of  $V_{PMT}$  is clearly visible in the model data (see Figure 2).



**Figure 2.** Change in measured intensity of Cy5-labeled cDNA spots of equal initial intensity with increasing number of scans, depending on the PMT voltage.

All in all, modelling of both parameters including  $V_{PMT}$  might yield a significant benefit in accuracy and it is therefore carried out and applied to our Cy5 model:

$$\ln(I(I_0, n_{scan}, V_{PMT})) = \ln(I_0) \times e^{(-\lambda(V_{PMT}) \times (n_{scan}-1)^{a(V_{PMT})})} \quad (6)$$

where  $I(I_0, n_{scan}, V_{PMT})$ : post PMT foreground intensity after  $n$  scans, with given  $I_0$  and  $V_{PMT}$ ;  $I_0$ : initial post PMT foreground intensity ( $n_{scan} = 1$ );  $n_{scan}$ : number of scans;  $V_{PMT}$ : PMT voltage;  $\lambda(V_{PMT})$ : degradation coefficient;  $a(V_{PMT})$ : exponent.

Both  $\lambda(V_{PMT})$  and  $a(V_{PMT})$  were modeled using second order polynomials. Based on the Cy5 model, a fit for Cy3 model data was calculated by varying  $\lambda$  and  $a$  for each  $V_{PMT}$  setting. The resulting parameters were examined using ANOVAs analogous to the Cy5 procedures, yielding comparable results. The  $V_{PMT}$  influence was then modeled using second order polynomials. The results are given in term (7) and (8) as well as table 3:

$$\lambda(V_{PMT}) = p_1 \times V_{PMT}^2 + p_2 \times V_{PMT} + p_3 \quad (7)$$

where  $\lambda(V_{PMT})$ : degradation coefficient;  $V_{PMT}$ : PMT voltage;  $p_1, p_2, p_3$ : paramters.

$$a(V_{PMT}) = p_1 \times V_{PMT}^2 + p_2 \times V_{PMT} + p_3 \quad (8)$$

where  $a(V_{PMT})$ : degradation exponent;  $V_{PMT}$ : PMT voltage;  $p_1, p_2, p_3$ : parameters.

**Table 3.** Parameters of the final fits.

Fluorophore	$\lambda(V_{PMT})$			$a(V_{PMT})$		
	$p_1$	$p_2$	$p_3$	$p_1$	$p_2$	$p_3$
Cy3	-2.153E-07	3.232E-04	-9.200E-02	1,106E-06	-1.885E-03	1.461
Cy5	-1.122E-08	1.640E-5	-1.948E-03	-4.533E-07	9.433E-05	0.901

#### 4.3. Model Analysis

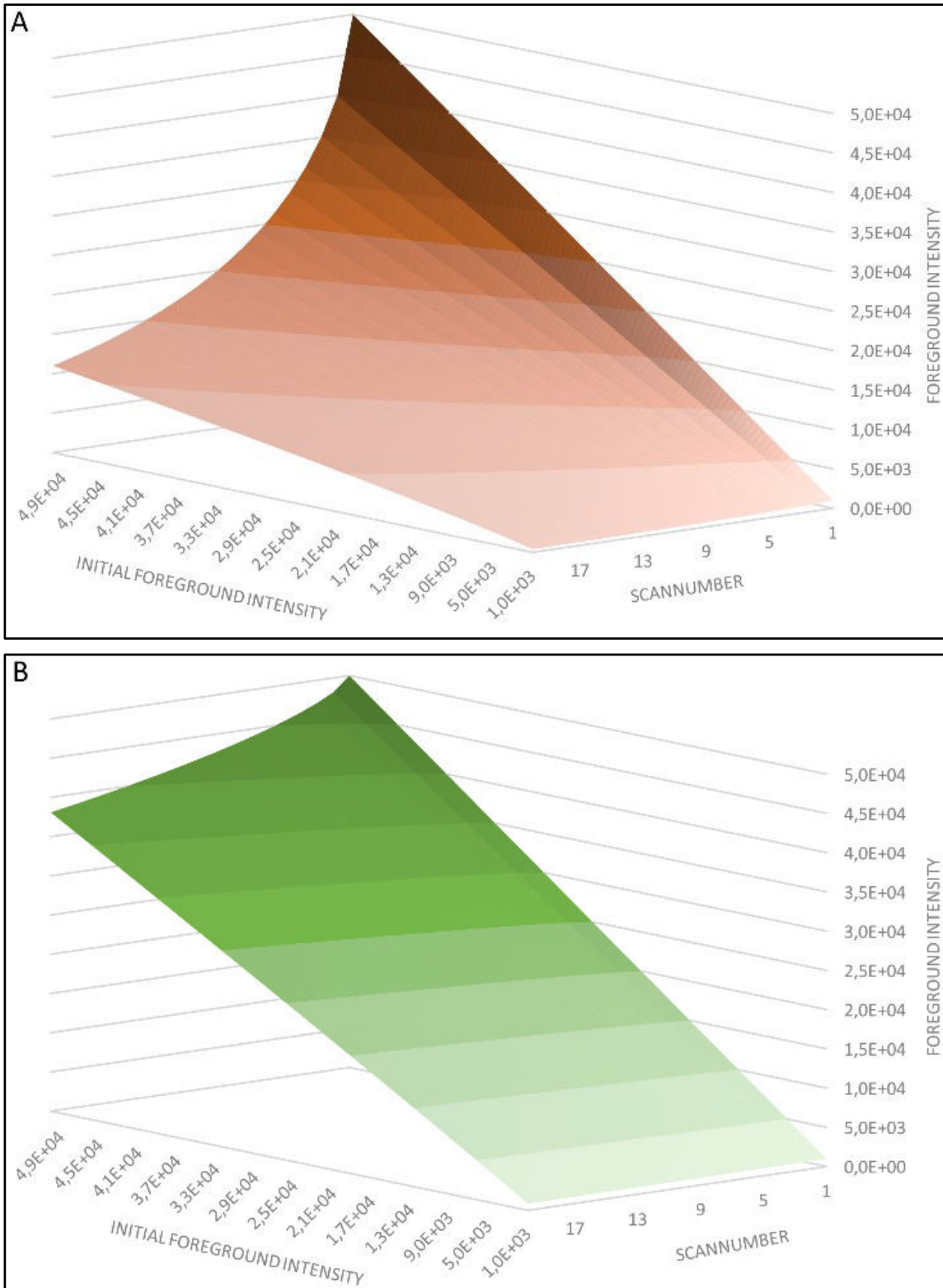
Both resulting models (shown in Figure 3) describe the observed bleaching effects to a high degree ( $R^2$  from 0.976 to 0.998 for different  $V_{PMT}$  settings, examples shown in Figure 4). The unequal susceptibilities of Cy3 and Cy5 to photobleaching clearly stand out: While Cy3-tagged spots lose between 23.19% and 32.01% of their observed intensity after 20 scans, the intensity of Cy5-tagged spots decrease between 76.92% and 87.07%. As can be seen in the model, the variance in signal decrease is introduced by the  $V_{PMT}$  settings, which shows that its incorporation into the model is crucial to remedy bias caused by bleaching. Looking at a scan number more likely to be utilized in daily microarray analysis, even after 5 scans the effect profoundly influences the observed intensities: Decreases of 8.73%–10.43% for Cy3 and 41.77%–52.97% for Cy5 emphasize the need for photobleaching correction and scanning protocol standardization not only for multiscan techniques, but for every application relying on microarray scan imaging. Furthermore, the dye-dependent bleaching-variation calls for a re-evaluation of dye swap and dye switch applications as well as mathematical tools designed to compensate for dye introduced bias (LOESS/LOWESS).

#### 4.4. First Model Validation

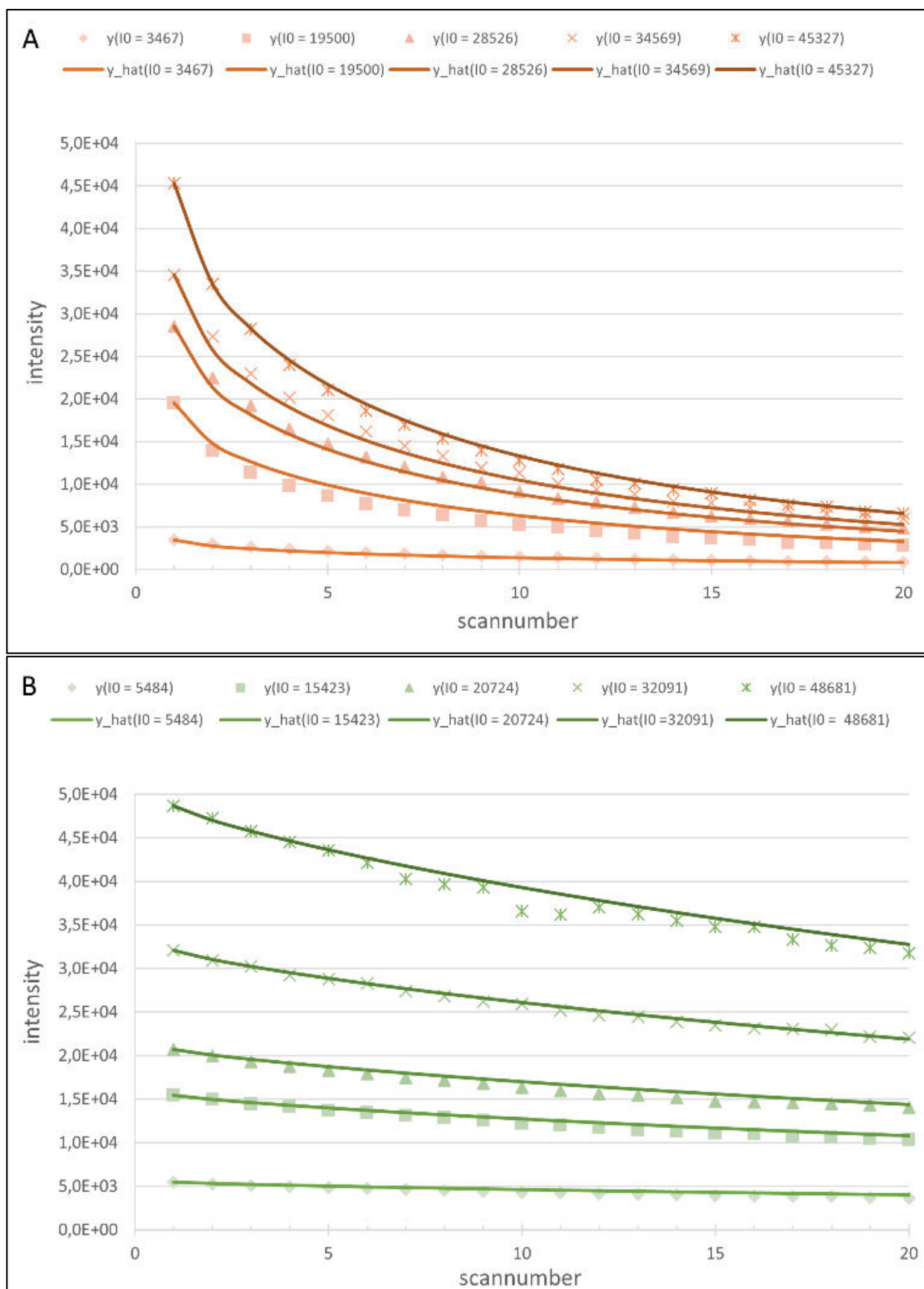
Following the model generation and characterization a model-based correction for photobleaching was carried out. The source data for this procedure (validation data) was recorded in a manner designed to emulate a random multiscan procedure. The slides used were manufactured analogous to their model data counterparts.

A basic principle of multiscan procedures lies in the correction of saturated or noisy spots through extrapolation of intensity data of different  $V_{PMT}$  settings. The reliability of the related extrapolation model is based on how well-defined its parameters are. In order to get a first assessment of the effect of photobleaching correction onto parameter quality, linear fits were calculated for data series of the same spots with differing  $V_{PMT}$ . Fits were calculated for each cyanine dye separately, with raw validation data and model corrected validation data. As seen in Table 4, the application of our model reduces the overall variability ( $\sigma_{\text{coefficient}}$ ,  $\sigma_{\text{intercept}}$ ), thereby improving the data's suitability for generating an extrapolation model ( $R^2$ ). The overall low coefficients of determination imply that a reasonable amount of variation remains. While the data was filtered in terms of noise and saturation, other source for variation were not addressed e.g., background intensity. No background correction was applied to the utilized data, as the background itself might be subject to photobleaching. This, and the ongoing discussion if the subtraction

of background intensity, is actually beneficial in terms of variability reduction [26] were the reasons for refraining from any background normalization. The characterization of the effect of photobleaching to the background will be the subject of future investigations.



**Figure 3.** (A) Three-dimensional illustration of the final model of Cy5-photobleaching for  $V_{PMT} = 950$  (B) Three-dimensional illustration of the final model of Cy3-photobleaching for  $V_{PMT} = 700$ .



**Figure 4.** (A) Cy5-data sets ( $y(I_0)$ ) with model-data ( $y\_hat(I_0)$ ) at  $V_{PMT} = 750$ .  $R^2$ : 0.994, 0.995, 0.998, 0.998, 0.999 (from lowest  $I_0$  to highest); and (B) Cy5-data sets ( $y(I_0)$ ) with fits ( $y\_hat(I_0)$ ) at  $V_{PMT} = 500$ .  $R^2$ : 0.994, 0.997, 0.991, 0.995, 0.984 (from lowest  $I_0$  to highest).

**Table 4.** Comparison of regression features of linear fit of  $\ln(I)$  vs.  $V_{PMT}$  for raw validation data and model corrected validation data for Cy5 and Cy3. Displayed are the mean  $R^2$  as well as the mean  $\sigma$  for both parameters of the linear fit for both cyanine dyes for uncorrected and corrected validation data.

Fluorophore	Regression Feature	Data Source	
		Raw Validation Data	Model Corrected Validation Data
Cy5	$\overline{R^2}$	<b>0.825</b>	<b>0.8384</b>
	$\overline{\sigma}_{\text{coefficient}}$	<b>44.222</b>	<b>26.429</b>
	$\overline{\sigma}_{\text{intercept}}$	<b><math>2.695 \times 10^4</math></b>	<b><math>1.120 \times 10^4</math></b>
Cy3	$\overline{R^2}$	<b>0.818</b>	<b>0.833</b>
	$\overline{\sigma}_{\text{coefficient}}$	<b>81.908</b>	<b>29.613</b>
	$\overline{\sigma}_{\text{intercept}}$	<b><math>5.038 \times 10^4</math></b>	<b><math>1.258 \times 10^4</math></b>

## 5. Conclusions

Our aim was to characterize and quantify the impact of photobleaching for DNA microarrays. Several groups have previously published approaches to improve the quality and capability of DNA microarray experiments, especially the extension of the linear range through multi-scan protocols constitutes a promising tool. We identified and characterized a major bias for multi-scan procedures and present a way to correct for this bias. In summary, we were able to generate models that explain photobleaching induced variability in multiscan microarray experiments for the two most commonly used fluorophore dyes, Cy3 and Cy5. Our models take into account the initial foreground intensity ( $I_0$ ), the number of carried out scans ( $n_{scan}$ ) as well as the current intensity ( $I$ ) recorded with a defined PMT voltage ( $V_{PMT}$ ). Parallel to the generation of these models we characterized the photobleaching effect of both abovementioned dyes, demonstrating the need for correction of this phenomenon not only for multiscan applications, but for all microarray scan based methods, e.g., our model, which explains the variability to a highly significant level and shows that the bleaching, itself, is not a simply linear subtractive effect. We therefore assume that a mere correction of the dye effect does not correct for the photobleaching by which the spots have been affected. A dye swap will in fact correct for intensity differences introduced by the choice of dye, but if the spots also differ in intensity, which they almost always will to a certain degree, photobleaching will not be automatically be co-corrected as it is not a linear additive effect. The degree of influence this effect has on microarray scans, and its disparity depending on the involved dye and the intensity level therefore calls for re-evaluation of dye swap/switch applications and dye effect normalization methods. As photobleaching is, to a lesser degree, induced by environmental light and other environmental factors, such as ozone concentration, our results suggest a standardization of microarray-slide handling to achieve comparable, if possible, minimal exposition to light prior to the scanning process. We are aware that a total lab-to-lab comparability in terms of microarray processing is not realistic, but still want to address the influence of environmental factors on bleaching and the overall quality of microarray results. A real standardization will not be accomplished by one single step, but through raising awareness of the subject we hope to help improve the reproducibility within a lab/workgroup. The benefit of correcting photobleaching-induced variability in multiscan applications was demonstrated. Corrected data was more suitable to generate linear  $\ln(I)$  vs.  $V_{PMT}$  fits, leading to more narrowly defined parameters. Future

studies need to validate these findings for actual hybridization experiments with dye-functionalized cDNA, accounting for the hybridization-derived effects on photobleaching involving the inclusion of the interdependent factors of intra-spot heterogeneity and FRET and non-FRET crosstalk. Several other factors need to be evaluated to apply our findings to DNA hybridization experiments in general. Among these the influence of temperature, DNA chain sequence and rigidity, dye concentration, and dye stacking. The overall physico-chemical characteristics of surface bound oligonucleotides are still to be sufficiently characterized [8,31]. Also the effect of photobleaching on background intensity needs to be examined to allow for integration of background correction. Likewise, interactions with other normalization methods have to be evaluated.

We encourage users of the technology to apply this information and develop multiscan solutions that correct for photobleaching.

### Author Contributions

Marcel von der Haar, Patrick Lindner and Frank Stahl conceived and designed the experiments. Marcel von der Haar and John-Alexander Preuß performed the experiments. Marcel von der Haar, John-Alexander Preuß, Kathrin von der Haar, Patrick Lindner, Thomas Scheper and Frank Stahl analyzed the data. Marcel von der Haar wrote the manuscript. All authors participated in the design of the study and approved the final manuscript.

### Conflicts of Interest

The authors declare no conflict of interest.

### References

1. Spielbauer, B.; Stahl, F. Impact of microarray technology in nutrition and food research. *Mol. Nutr. Food Res.* **2005**, *49*, 908–917.
2. Allison, D.B.; Cui, X.; Page, G.P.; Sabripour, M. Microarray data analysis: From disarray to consolidation and consensus. *Nat. Rev. Genet.* **2006**, *7*, 55–65.
3. Ehrenreich, A. DNA microarray technology for the microbiologist: An overview. *Appl. Microbiol. Biotechnol.* **2006**, *73*, 255–273.
4. Kretschy, N.; Somoza, M.M. Comparison of the sequence-dependent fluorescence of the cyanine dyes cy3, cy5, dylight dy547 and dylight dy647 on single-stranded DNA. *PLoS ONE* **2014**, *9*, e85605.
5. Mary-Huard, T.; Daudin, J.J.; Robin, S.; Bitton, F.; Cabannes, E.; Hilson, P. Spotting effect in microarray experiments. *BMC Bioinf.* **2004**, doi:10.1186/1471-2105-5-63.
6. Dawson, E.D.; Reppert, A.E.; Rowlen, K.L.; Kuck, L.R. Spotting optimization for oligo microarrays on aldehyde-glass. *Anal. Biochem.* **2005**, *341*, 352–360.
7. Sobek, J.; Aquino, C.; Weigel, W.; Schlapbach, R. Drop drying on surfaces determines chemical reactivity-the specific case of immobilization of oligonucleotides on microarrays. *BMC Biophys.* **2013**, doi:10.1186/2046-1682-6-8.
8. Rao, A.N.; Grainger, D.W. Biophysical properties of nucleic acids at surfaces relevant to microarray performance. *Biomater. Sci.* **2014**, *2*, 436–471.

9. Jang, H.; Cho, M.; Kim, H.; Kim, C.; Park, H. Quality control probes for spot-uniformity and quantitative analysis of oligonucleotide array. *J. Microbiol. Biotechnol.* **2009**, *19*, 658–665.
10. Khondoker, M.R.; Glasbey, C.A.; Worton, B.J. Statistical estimation of gene expression using multiple laser scans of microarrays. *Bioinformatics* **2006**, *22*, 215–219.
11. Satterfield, M.B.; Lipka, K.; Lu, Z.Q.; Salit, M.L. Microarray scanner performance over a five-week period as measured with cy5 and cy3 serial dilution slides. *J. Res. Natl. Inst. Stand. Technol.* **2008**, *113*, 157–174.
12. Ambroise, J.; Bearzatto, B.; Robert, A.; Macq, B.; Gala, J.L. Combining multiple laser scans of spotted microarrays by means of a two-way anova model. *Statist. Appl. Genet. Mol. Biol.* **2012**, doi:10.1515/1544-6115.1738.
13. Vora, G.J.; Meador, C.E.; Anderson, G.P.; Taitt, C.R. Comparison of detection and signal amplification methods for DNA microarrays. *Mol. Cell. Probes* **2008**, *22*, 294–300.
14. Shi, L.; Tong, W.; Su, Z.; Han, T.; Han, J.; Puri, R.K.; Fang, H.; Frueh, F.W.; Goodsaid, F.M.; Guo, L.; *et al.* Microarray scanner calibration curves: Characteristics and implications. *BMC Bioinf.* **2005**, doi:10.1186/1471-2105-6-S2-S11.
15. Lyng, H.; Badiee, A.; Svendsrud, D.H.; Hovig, E.; Myklebost, O.; Stokke, T. Profound influence of microarray scanner characteristics on gene expression ratios: Analysis and procedure for correction. *BMC Genomics* **2004**, doi:10.1186/1471-2164-5-10.
16. Dar, M.; Giesler, T.; Richardson, R.; Cai, C.; Cooper, M.; Lavasani, S.; Kille, P.; Voet, T.; Vermeesch, J. Development of a novel ozone- and photo-stable hyper5 red fluorescent dye for array cgh and microarray gene expression analysis with consistent performance irrespective of environmental conditions. *BMC Biotechnol.* **2008**, doi:10.1186/1472-6750-8-86.
17. Kuang, C.; Luo, D.; Liu, X.; Wang, G. Study on factors enhancing photobleaching effect of fluorescent dye. *Measurement* **2013**, *46*, 1393–1398.
18. Drăghici, S. *Statistics and Data Analysis for Microarrays Using R and Bioconductor*, 2nd ed.; Taylor & Francis: Boca Raton, FL, USA, 2011.
19. Dobbin, K.; Shih, J.H.; Simon, R. Statistical design of reverse dye microarrays. *Bioinformatics* **2003**, *19*, 803–810.
20. Gupta, R.; Auvinen, P.; Thomas, A.; Arjas, E. Bayesian hierarchical model for correcting signal saturation in microarrays using pixel intensities. *Statist. Appl. Genet. Mol. Biol.* **2006**, doi:10.1155/2008/231950.
21. Bengtsson, H.; Jonsson, G.; Vallon-Christersson, J. Calibration and assessment of channel-specific biases in microarray data with extended dynamical range. *BMC Bioinf.* **2004**, doi:10.1186/1471-2105-5-177.
22. Williams, A.; Thomson, E.M. Effects of scanning sensitivity and multiple scan algorithms on microarray data quality. *BMC Bioinf.* **2010**, doi:10.1186/1471-2105-11-127.
23. Kerr, M.K.; Afshari, C.A.; Bennett, L.; Bushel, P.; Martinez, J.; Walker, N.J.; Churchill, G.A. Statistical analysis of a gene expression microarray experiment with replication. *Stat. Sin.* **2002**, *12*, 203–217.
24. Stennett, E.M.S.; Ciuba, M.A.; Levitus, M. Photophysical processes in single molecule organic fluorescent probes. *Chem. Soc. Rev.* **2014**, *43*, 1057–1075.



25. Staal, Y.C.M.; van Herwijnen, M.H.M.; van Schooten, F.J.; van Delft, J.H.M. Application of four dyes in gene expression analyses by microarrays. *BMC Genomics* **2005**, doi:10.1186/1471-2164-6-101.
26. Qin, L.X.; Kerr, K.F. Empirical evaluation of data transformations and ranking statistics for microarray analysis. *Nucleic Acids Res.* **2004**, *32*, 5471–5479.
27. Iqbal, A.; Arslan, S.; Okumus, B.; Wilson, T.J.; Giraud, G.; Norman, D.G.; Ha, T.; Lilley, D.M.J. Orientation dependence in fluorescent energy transfer between cy3 and cy5 terminally attached to double-stranded nucleic acids. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 11176–11181.
28. Ha, T.; Enderle, T.; Ogletree, D.F.; Chemla, D.S.; Selvin, P.R.; Weiss, S. Probing the interaction between two single molecules: Fluorescence resonance energy transfer between a single donor and a single acceptor. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 6264–6268.
29. Rao, A.N.; Rodesch, C.K.; Grainger, D.W. Real-time fluorescent image analysis of DNA spot hybridization kinetics to assess microarray spot heterogeneity. *Anal. Chem.* **2012**, *84*, 9379–9387.
30. Rao, A.N.; Vandencastele, N.; Gamble, L.J.; Grainger, D.W. High-resolution epifluorescence and time-of-flight secondary ion mass spectrometry chemical imaging comparisons of single DNA microarray spots. *Anal. Chem.* **2012**, *84*, 10628–10636.
31. Harrison, A.; Binder, H.; Buhot, A.; Burden, C.J.; Carlon, E.; Gibas, C.; Gamble, L.J.; Halperin, A.; Hooyberghs, J.; Kreil, D.P.; *et al.* Physico-chemical foundations underpinning microarray and next-generation sequencing experiments. *Nucleic Acids Res.* **2013**, *41*, 2779–2796.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).

### 3.2 Optimization of Cyanine Dye Stability and Analysis of FRET Interaction on DNA Microarrays

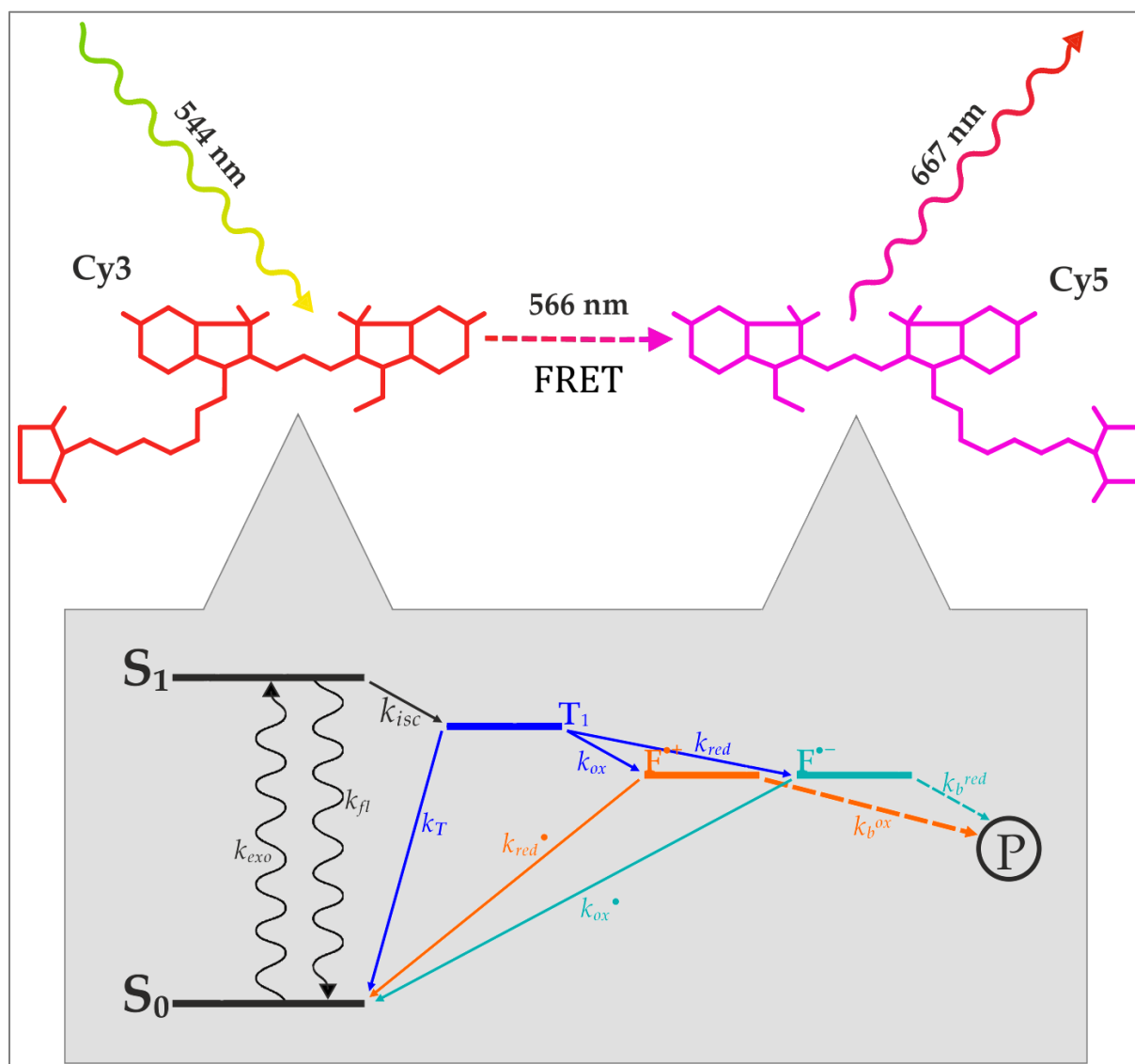


Figure 3.2: Graphical abstract of „Optimization of Cyanine Dye Stability and Analysis of FRET Interaction on DNA Microarrays”

In the previous section, the impact of photobleaching on cyanine dye labeled DNA microarray setups was characterized. The need for correction of this bias, particularly for multi-scan design was addressed by the development of an empirical model suited to normalize occurring bleaching-bias. A major advantage of this model is the possibility to apply it to microarray data as is, without the need of changing the experimental design or using commercial, modified dyes. It can easily be applied to previously generated data too. The research and development reported upon in this section were carried out with the same paradigm in mind: Expanding the knowledge and understanding of underlying processes to develop solutions that can be easily applied by the

experimenter. Following the conviction that the worth of a new method is also determined by its practical applicability.

Further analysis of the data generated in chapter 3.1 unveiled a phenomenon observable in spots where Cy3 and Cy5 labeled oligos were present. In these spots, an increase of Cy3 intensity was observed. This behavior – previously been described as “passive de-quenching” – is caused by a combination of photobleaching and FRET [18,34,35,67]. While functional Cy5 molecule are in proximal vicinity to Cy3, some Cy3 molecule transfer the energy of their excited electrons to Cy5 molecules by FRET. This causes a smaller observable intensity of Cy3. Over the course of multiple scans, an increasing amount of Cy5 molecule became unavailable for FRET as they were subjected to photo destruction. While Cy3 molecule were – to a lesser degree – also affected by bleaching, the chances for these cyanine molecule to emit photons themselves increased, as the probability of having photoactive Cy5 that could act as a FRET acceptor in their vicinity decreased. These processes resulted in a higher observable Cy3-related intensity in spite of the influence of photobleaching.

As these processes are not accounted for by the model from chapter 3.1, an alternative approach was devised. Based on works of Vogelsang *et al.* a protective buffer was manufactured [20]. This buffer contains a ROXS that catalyzes the depopulation of those intermediate states of excited cyanine electrons from which photo destruction can be initialized. This ROXS buffer is used as part of an additional functional layer mounted onto the microarray slide prior to the scanning process. This setup allows the scanning of microarray slide while the cyanine dye are environed by a medium, protecting them from ozone and photo induced bleaching. The simple construction and compatibility with the widespread GenePix® 4000B Microarray Scanner by Molecular Devices (Sunnyvale, CA, USA) of this solution again allows for easy applicability for all laboratories working with DNA microarray technology.

The significant protection against photobleaching that this method provides also allowed for the characterization of FRET in DNA microarray experiments. FRET became observable when comparing intensity changes of ROXS protected vs. unprotected arrays over the course of multiple scans. It can be shown that the utilization of the same ROXS buffer system is also suited as a means to normalizing FRET bias, further increasing the applicability and comparability of microarray gene expression data.

Article

# Optimization of Cyanine Dye Stability and Analysis of FRET Interaction on DNA Microarrays

Marcel von der Haar \*, Christopher Heuer, Martin Pähler, Kathrin von der Haar, Patrick Lindner, Thomas Scheper and Frank Stahl

Institute of Technical Chemistry, Leibniz University Hanover, Callinstr. 5, 30167 Hanover, Germany; christopher.heuer@gmx.net (C.H.); paehler@iftc.uni-hannover.de (M.P.); vonderhaar@iftc.uni-hannover.de (K.v.d.H.); lindner@iftc.uni-hannover.de (P.L.); scheper@iftc.uni-hannover.de (T.S.); stahl@iftc.uni-hannover.de (F.S.)

\* Correspondence: koch@iftc.uni-hannover.de; Tel.: + 49-511-762-2316

Academic Editor: Chris O'Callaghan

Received: 29 September 2016; Accepted: 24 November 2016; Published: 30 November 2016

**Abstract:** The application of DNA microarrays for high throughput analysis of genetic regulation is often limited by the fluorophores used as markers. The implementation of multi-scan techniques is limited by the fluorophores' susceptibility to photobleaching when exposed to the scanner laser light. This paper presents combined mechanical and chemical strategies which enhance the photostability of cyanine 3 and cyanine 5 as part of solid state DNA microarrays. These strategies are based on scanning the microarrays while the hybridized DNA is still in an aqueous solution with the presence of a reductive/oxidative system (ROXS). Furthermore, the experimental setup allows for the analysis and eventual normalization of Förster-resonance-energy-transfer (FRET) interaction of cyanine-3/cyanine-5 dye combinations on the microarray. These findings constitute a step towards standardization of microarray experiments and analysis and may help to increase the comparability of microarray experiment results between labs.

**Keywords:** microarray; DNA; scanning; photobleaching; fluorophore; cyanine dye; FRET; ROXS; bioinformatics; bioanalytics

## 1. Introduction

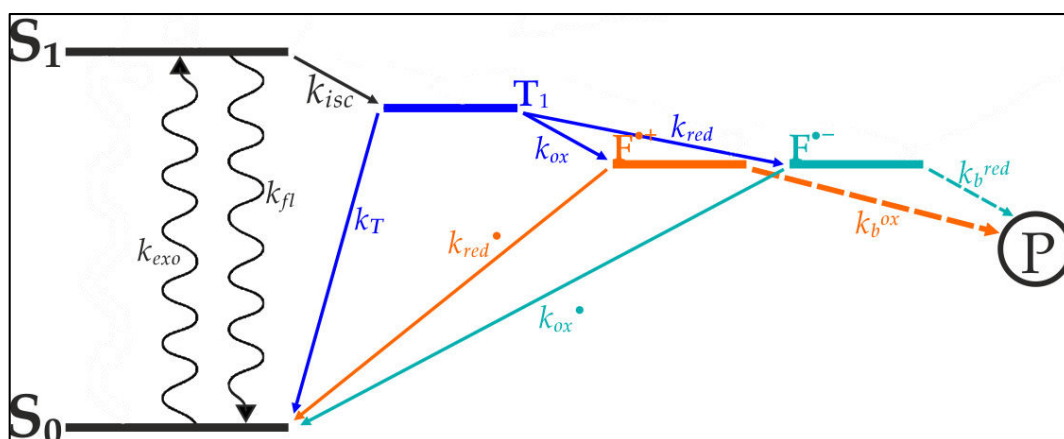
DNA microarrays are a potent technology for high throughput gene regulation monitoring. Fluorescence-labeled complementary DNAs (cDNAs) are transcribed from mRNA which is acquired from different regulatory states of the chosen biological sample. These cDNAs are competitively hybridized on a modified glass slide. The differently labeled fluorophore cDNA-probes compete for binding spotted, immobilized DNA-targets. The ratio of the differently labeled, immobilized fluorophores on a spot therefore represents the relative abundance of RNA in the respective regulatory states. Technology based upon this principle has gained widespread use in molecular biology, genetics, and medicine [1,2], enabling high-throughput transcriptome analysis [3].

Nonetheless, DNA microarray technology is set back by a set of disruptive factors, limiting its application and potential exploitation. These technical, biochemical, and statistical biases are introduced in various steps of a DNA microarray experiment. Sequence-dependent bias is introduced by primer design [4], spot-geometry and homogeneity through choice of spotting technique, proximate humidity, and choice of buffer [5–9]. Further bias is introduced by choice of dyes, scanner settings, the presence/absence of ozone filters, the exposition to environmental light, etc. [10–17]. While significant results can still be acquired in spite of these bias sources, they still pose a substantial hindrance when it comes to lab to lab comparability and standardization [18].

This publication focuses on photonic and photochemical effects, such as photobleaching and energy transfer, that emerge while scanning the DNA microarrays. In previous works, photobleaching

susceptibility of the almost omnipresent labeling agents cyanine-3 (Cy3) and cyanine-5 (Cy5) on DNA microarrays was investigated. Its effect on scanner data was successfully characterized and an empirical model was devised. The model efficiently normalizes the bias introduced by this effect with respect to the choice of dye, the previously carried out scans, and the scanner settings [19]. This study aims to minimize photobleaching of Cy3 and Cy5 using a reductive-oxidative, protective buffer (ROXS).

Vogelsang et al. [20] were able to show that blinking and photo destruction of cyanine dyes could be significantly reduced through depopulation of reactive intermediate states of the cyanine's excited electrons [20]. As seen in Figure 1, internal transitions of the cyanine's excited electron lead to a triplet-state from which photo destruction originates. The depletion of this state minimizes the electrons availability for bleaching processes, thereby increasing the dye's longevity and quantum yield. Based on the works of Widengren et al., Vogelsang et al. designed a buffer which contains an oxidizing agent and a reducing agent [21]. These agents are aimed at catalyzing the transition from the triplet-state towards the ground state (Figure 1).

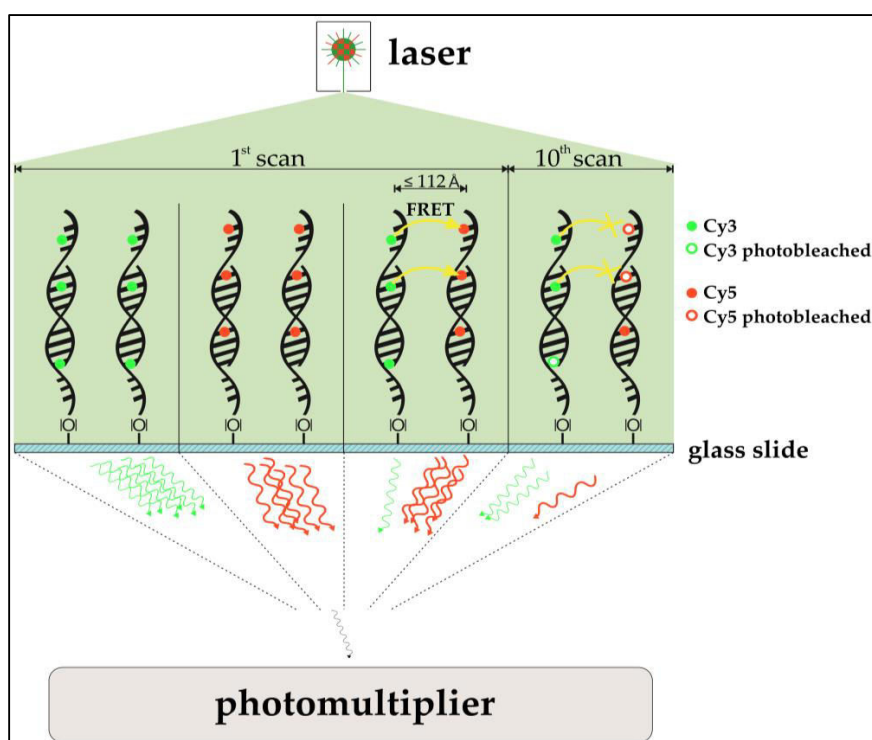


**Figure 1.** Schematic model of photoinduced electron excitation of organic fluorophores such as cyanine dyes. From its ground state ( $S_0$ ) the electron is excited to the first singlet state ( $S_1$ ). The sporadically forming triplet state ( $T_1$ ) is the point of origin for several transitions resulting in the formation of the photobleaching product (P). Methylviologen and ascorbic acid both rapidly deplete the  $T_1$ , forming a radical cation ( $F^{\bullet+}$  through methylviologen) or a radical anion ( $F^{\bullet-}$  through ascorbic acid). These radical ions rapidly recover through reduction (ascorbic acid) or oxidation (methylviologen). The combination of an oxidizing agent and a reducing agent (ROXS) therefore minimizes photoinduced formation of P. Model, according to Vogelsang et al. [20].

Vogelsang et al. [20] and others [22] carried out their experiments in aqueous solution using a fluorescence microscope. This study aims to apply their findings in DNA microarray experiments, where the cyanine dye is bound to the DNA which itself is fixated on a modified glass slide and is scanned using a microarray slide scanner. This change in experimental design necessitated an adapted approach on array design and scanning technique. In order to allow for scanning in the presence of the protective ROXS-buffer, the arrays were partially modified by adding an improvised liquid chamber (see Section 2).

In addition, the possible occurrence of Förster-resonance-energy-transfer (FRET, also Fluorescence-Energy-Resonance-Transfer) in DNA microarrays and its implications on microarray analysis were examined. FRET between Cy3 and Cy5 molecules has already been described and is a common tool for oligonucleotide analysis [23]. Although commercial alternatives to the aforementioned cyanine dyes exist, Cy3 and Cy5 are still used ubiquitously, and the need for optimizing these dyes' handling is compulsory.

The use of acceptor photobleaching (Cy5 being the acceptor) as a means to FRET validation is especially of interest for this study. In prior studies, the passive “de-quenching” of Cy3 by photo destruction of Cy5 resulted in an increase of donor (Cy3) photon emission, which was used to quantify FRET [24,25]. Among other findings, Rao et al. [26] qualitatively assessed if FRET is observable in DNA microarray two-dye experiments. To do so, a part of a spot containing Cy3- and Cy5-functionalized immobilized oligonucleotides was exposed to a confocal laser bleaching Cy5. Emissions of Cy3 and Cy5 were compared prior to and after the selective bleaching. In fact, the expected anti-proportional change in intensity for both dyes was observed, indicating that DNA microarray imaging of two-color experiments is biased by FRET. An investigation of the actual impact of FRET for multi-scan approaches is the subject of this study. Also, cross-over effects of FRET and ROXS protection are examined. A simplified model of the expected impact of the FRET-effect can be seen in Figure 2.



**Figure 2.** Schematic model of a hypothesized Förster-resonance-energy-transfer (FRET)-effect in cyanine-labeled two-dye DNA microarray scanning. Cyanine-labeled DNA in single-dye setups emits photons after excitation (upper left, upper mid left). In a two-dye setup at the first scan (upper mid right), Cy3 only partially emits photons after excitation. With Cy5 in its vicinity, Cy3 acts as a FRET donor, transferring the energy to the Cy5-acceptor, which itself emits a photon. This leads to a lower Cy3 signal and a higher Cy5 signal compared to Cy3 and Cy5 from single dye setups. After several scans, photobleaching should have decreased the amount of functional, photon-emitting cyanine molecules. While this would be observable in single dye setups (not shown), one does actually observe a different behavior in two-dye setups (upper left). The higher bleaching susceptibility of Cy5 decreased the chance of Cy3 acting as a FRET-donor, simultaneously increasing the amount of emitted photons from Cy3. While a strong decrease in Cy5-photon emission can be observed, the emission of Cy3 seems to have “increased”. This effect is called passive “de-quenching”. All emitted photons then enter the photomultiplier (PMT), where they are transformed into an exponentially enhanced electron signal.

Other than to improve the awareness and understanding of underlying photochemical processes and their effect on microarray data, the results of these studies are aimed at the improvement of microarray bias minimization and the establishment of experiment reproducibility and lab-to-lab comparability.

## 2. Materials and Methods

**Oligo Preparation:** Single strand DNAs (ssDNA) of 50 nt length were purchased from Integrated DNA Technologies, Inc. (Munich, Germany). The sequences were optimized with regard to low stabilities of potential homodimers and hairpins. The 3'-end of the ssDNA was modified with an amino-modified C6 spacer. Ninety-six different sequences were used, corresponding to a set of 96 *Escherichia coli* genes. This set of genes was chosen because it provides a representative set of regulatory behaviors for heat-shock experiments. Also, the usage and analysis of these genes is well documented and routinely carried out in our workgroup. Information on these genes can be found in the Supplementary Materials (Table S1). The oligos were dissolved using Micro Spotting Solution Plus 2X from Arrayit Corporation (Sunnyvale, CA, USA) and nuclease free water to a final concentration of 100 mM (concentrations validated using a NanoDrop 2000 from Thermo Fisher Scientific Inc. (Waltham, MA, USA)). Solutions were stored at 4 °C.

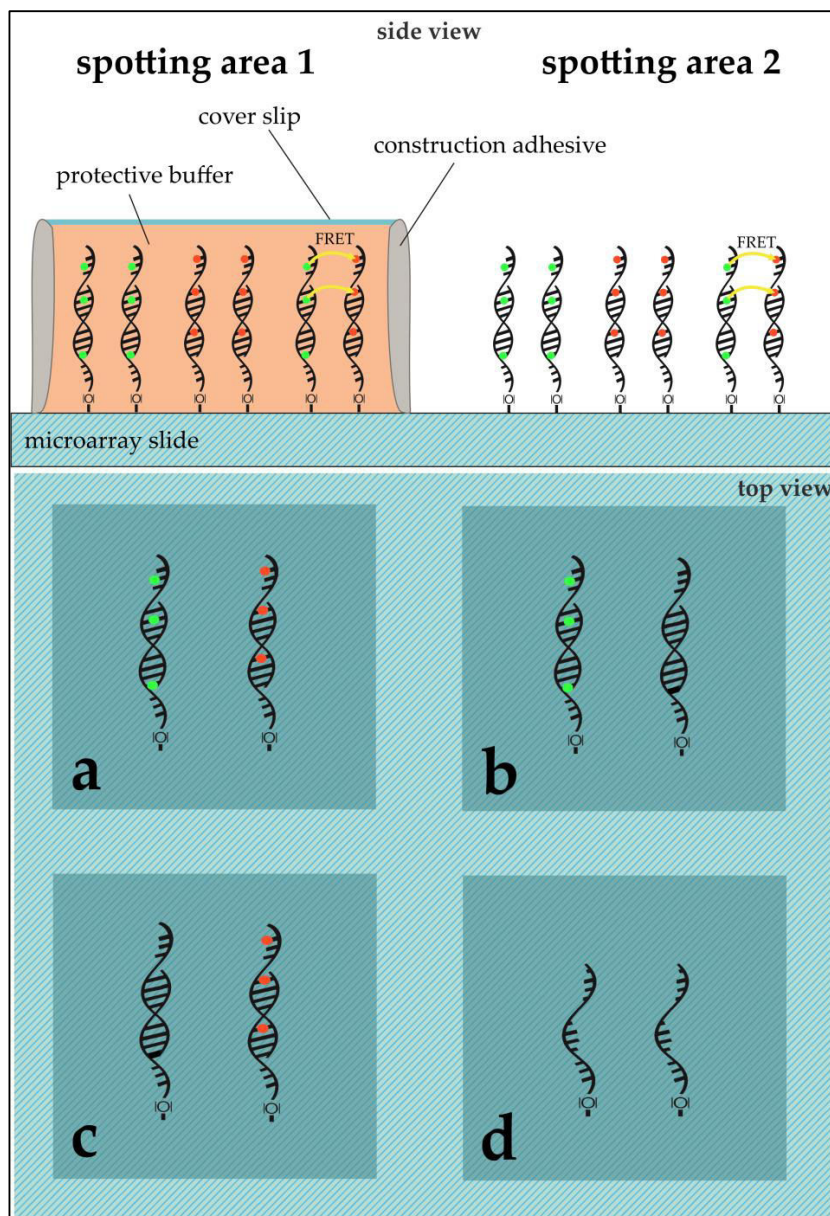
**ROXS Buffer Preparation:** ROXS buffers were prepared freshly prior to each experiment. They were based on a 1× standard buffer from phosphate buffered saline (PBS) at pH 7.4, containing additional ascorbic acid (AA) and methylviologen dichloride hydrate (MV) at 100 mM each. Dilutions of this stock solution were prepared using 1× PBS. Consequently, if a buffer is described as, for example, 10 mM ROXS, it contains 10 mM AA and 10 mM MV in 1× PBS.

**DNA Immobilization:** DNA sequences were immobilized on aldehyde modified glass slides (SuperAldehyde 2; Arrayit<sup>®</sup> Corporation, Sunnyvale, CA, USA) using a non-contact-spotter (Nano Plotter<sup>™</sup> NP2.1; GeSiM mbH, Großerkmannsdorf, Germany) with an applied voltage of 80 V. The selection of a contact-free printer allowed for higher homogeneity in spot geometry by avoiding pin-derived variance and providing humidity control in the spotting chamber (humidity at 60%). The general spotting layout can be found in Figure 3.

**RNA Treatment and On-Slide Hybridization:** RNA was purified and pooled from samples of two different treatments using Trizol reagent (Invitrogen, Karlsruhe, Germany) according to the manufacturer's protocol. This method yielded an average of 30 µg total RNA from 10<sup>6</sup> cells. In both cases, *E. coli* was cultivated until it reached the log-phase at 37 °C. While the 37 °C sample (Ec37) was obtained in this phase directly, the 50 °C sample (Ec50) *E. coli* was exposed to 50 °C for ten minutes before cell disruption and RNA purification. Fifty micromoles of purified DNA was transcribed into complementary DNA (cDNA) using a 1:1:1:1 unlabeled dNTP-mixture for unlabeled cDNA and a 1:1:1:0.25 unlabeled dNTP-mixture (with dCTP being the aforementioned 0.25) with the addition of 0.75 equivalents of Cy3- or Cy5-labeled dCTPs. In the case of labeling, Cy3 was always used for Ec37 while Ec50 was labelled with Cy5. The purified cDNAs were then competitively hybridized on the microarray slides. The hybridized microarray slides were put into cassettes, purchased from Arrayit Corporation, for microarray sample multiplexing. Sixteen microliters of the desired cDNA solution was pipetted into the wells (see Figure 3). The cassette's wells were sealed using an adhesive strip to prevent dehydration and the arrays were hybridized at 100% humidity overnight. The slides were washed and dried through centrifugation.

**Application of PBS and/or ROXS-Buffer:** The spotting pattern allowed for two different treatments per slide. Possible treatments were: unprotected (bare slide, without any protection) or 1× PBS/10 mM ROXS/50 mM ROXS (40 µL of buffer were pipetted onto the slide, covered with a cover slip that was sealed using construction adhesive).

**Microarray Slide Scanning:** All scans were performed using the GenePix<sup>®</sup> 4000B Microarray Scanner by Molecular Devices (Sunnyvale, CA, USA). All data was collected at a pixel size of 10 µm and a total resolution of 1891 × 2089 pixels. For unprotected areas the focus level remained by default at 0 µm. Areas protected by a cover slip were scanned at a focus level of 75 µm.



**Figure 3.** Microarray modified glass slide scheme for ROXS and FRET assessment. The slide shows two main spotting areas (1,2), each subdivided into four blocks (a–d). Each block was used to immobilize either 96 capture-oligos without replication or 24 capture-oligos with five replicates (six spots per gene). While (a) was used to hybridize Cy3- and Cy5-labeled cDNAs competitively; (b) for Cy3 vs. unlabeled cDNA; (c) for unlabeled vs. Cy5; and (d) was used as a negative control, where no hybridization took place. In the case of subsequent application of a protective cover slip, area 1 remained unprotected while area 2 was modified using a desired buffer and a cover slip (see **Application of PBS and/or ROXS-Buffer**).

The usage of a different focus level for areas modified with a liquid film and a cover slip was imperative to maintain comparable imaging results. Each area was pre-scanned once to determine the scan-area and 10 additional scans of this area were performed at constant photomultiplier (PMT) voltages (635 nm-laser: 800 V, 532 nm-laser: 650 V) and 100% laser power. Data collection was carried out using GenePix<sup>®</sup> Pro 7.0 (Molecular Devices, Sunnyvale, CA, USA).



**Data Analysis:** After an initial quality control carried out by GenePix<sup>®</sup>Pro, all spots with any saturated pixels, as well as spots whose signal to noise ratio (SNR) was 3 or lower, were excluded from further analysis. The SNR is defined as follows:

$$SNR = \frac{m_{Foreground} - m_{Background}}{s_{Background}} \quad (1)$$

where  $m$ : median;  $s$ : standard deviation.

Also, in accordance with Lyng et al.'s recommendations [15], spots with median local background subtracted intensities above 50,000 and below 1000 relative intensity units were excluded from further analysis to prevent saturation and/or noise bias. Although a correction for background is a general convention, the actual application varies. Background correction is carried out locally, within a sub-grid, with blank spots or control spots. Most of these approaches have different underlying assumptions on how the background intensity reflects an intensity bias over- or better underlying the feature intensity. Furthermore, Qin et al. [27] showed that while a background subtraction actually reduces the bias it increases data variability. The increase in variability is kept in check using the SNR threshold. Data was filtered and analyzed using MATLAB v7.12.0.635 (The MathWorks, Inc., Natick, MA, USA) and Visual Basic for Applications (Microsoft Corp., Redmond, WA, USA).

All derived statistical metadata in this study was calculated taking into account Gaussian error propagation. If not stated otherwise, all error indicators given in text and graphs always represent the respective value's standard error of the mean (SEM) with a confidence level of 95.4%.

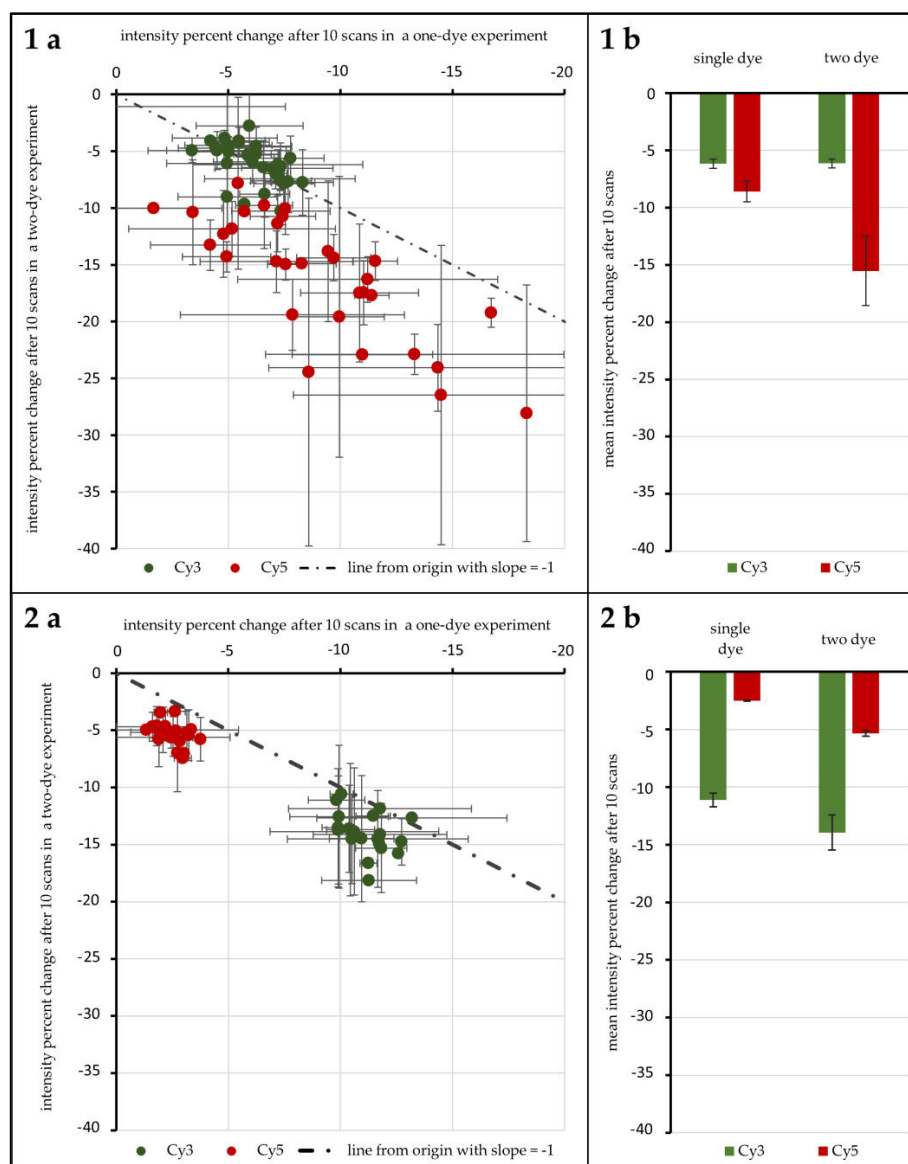
### 3. Results

#### 3.1. Ninety-Six Gene Experiment

A first test with 96 genes was carried out to test the influence of ROXS as well as FRET and possible cross-over effects. As Figure 4(1a,b) shows, without any protective measures, spots hybridized with Cy3-labeled DNA lose 6.16% ( $\pm 0.40\%$ ) in signal intensity on average after 10 scans when no Cy5-labeled DNA is present. Furthermore, 6.14% ( $\pm 0.38\%$ ) are lost when Cy3-labeled DNA is hybridized competitively against Cy5-labeled DNA. T-Tests show that these two values cannot be considered different ( $\alpha \leq 0.1$ ). For spots hybridized with Cy5-labeled DNA, on the other hand, a statistical difference is evident: hybridized against unlabeled DNA, Cy5-labeled DNA loses 8.61% ( $\pm 0.90\%$ ) on average. When hybridized against Cy3-labeled DNA, the intensity decrease changes to 15.52% ( $\pm 3.02\%$ ). These two means are significantly different for  $\alpha \leq 0.01$ . Statistical inquiries showed that the percentage intensity change evaluated here is independent of the initial intensity level of the spots under study, ruling out a possible intensity level bias (supporting data can be found in the Supplementary Materials, Tables S2 to S4). The use of protective measures (in this case 1 mM ROXS in  $1 \times$  PBS) was evaluated in comparison (see Figure 4(2a,b)). Here, Cy3 without Cy5 loses 11.14% ( $\pm 0.60\%$ ), compared to 13.95% ( $\pm 0.15\%$ ). This difference is as significant ( $\alpha \leq 0.01$ ) as the change of single Cy5 with a loss of 2.51% ( $\pm 0.05\%$ ) to Cy5 with Cy3 present, losing 5.34% ( $\pm 0.24\%$ ).

#### 3.2. Twenty-Four Gene Experiment

The 96 genes of the first experiment were spotted with one replica. The respective gene was analyzed only when sufficient data for all dye-combinations and treatments was available. The use of only one replica limited the amount of usable data (data from 61 genes could not be used because for at least one dye-combination/treatment only one spot met the quality criteria). In order to provide a more sufficient statistical basis to validate the 96 gene experiment and answer the remaining questions, a new experiment design was devised: from those 96 genes, 24 were selected that had the most stable and homogenous spots, also providing a signal variety concerning overall intensity level and Cy3/Cy5-intensity-ratios. The corresponding 24 oligos were spotted with five replications, providing a solid basis for statistical analysis.

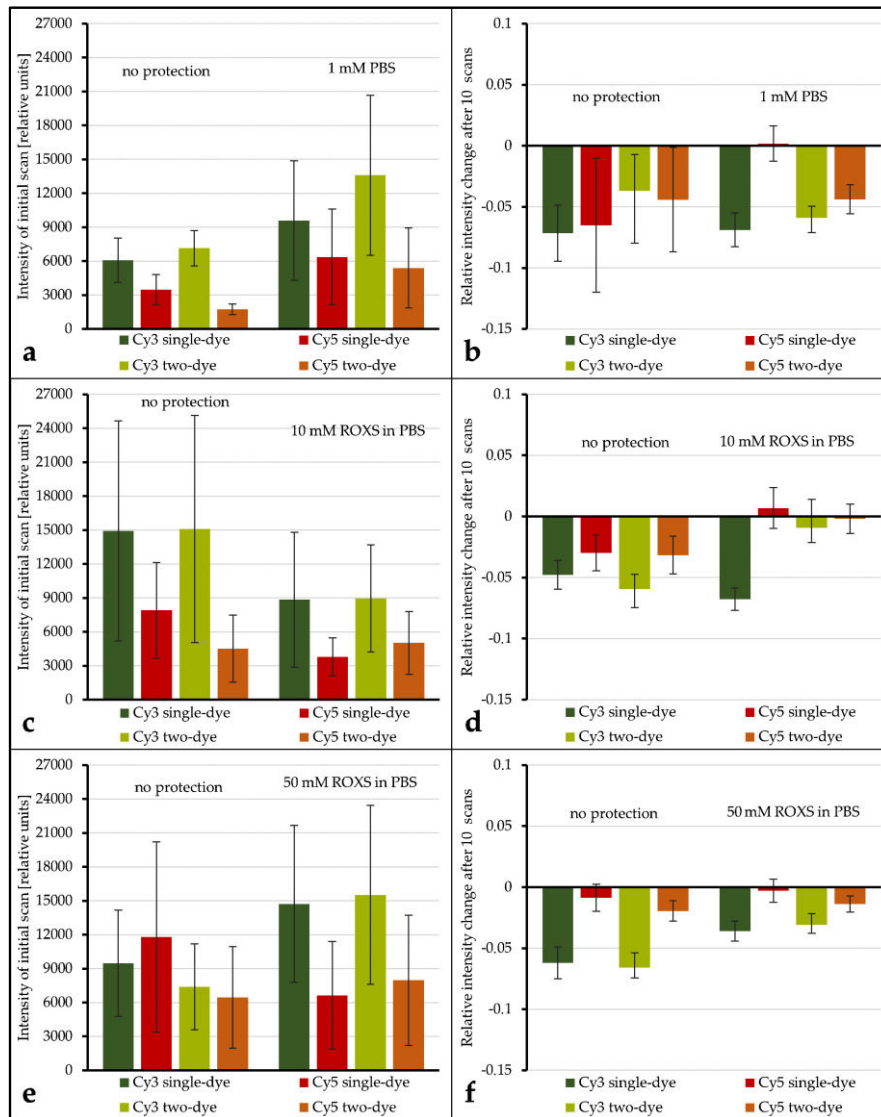


**Figure 4.** Results of the 96 gene experiment. The intensity percent change after 10 scans is compared for Cy3 and Cy5 depending on the presence/absence of their FRET partner. The percent change for one-dye setups is plotted against the same value derived from two dye experiments for unprotected spots (**1a**) and spots protected by 1 mM ROXS in 1× PBS (phosphate buffered saline) (**2a**). The dotted lines pass through the origin with a slope of  $-1$ . For each distribution the resulting means are given in (**1b,2b**). Error indicators are the respective standard errors (confidence: 95.4%) with Gaussian error propagation.

In addition to the evaluation of unprotected areas, several protective measures were compared: liquid chamber with 1× PBS, liquid chamber with 10 mM ROXS, and liquid chamber with 50 mM ROXS. With respect to array-to-array variability, each slide held one unprotected area and one protected area, to allow for array-to-array comparisons via normalization of the unprotected areas, resulting in the following pattern: Array1: unprotected vs. 1× PBS, Array2: unprotected vs. 10 mM ROXS in PBS, Array3: unprotected vs. 50 mM ROXS in PBS.

In a first evaluation of the results it could be confirmed that the three arrays' unprotected spots are statistically comparable with respect to intensity level, percentage intensity change, and overall spot intensity standard deviation (Tables 1 and 2, Supplementary Materials ANOVAs, Tables S2 to S4).

Even if the application of a liquid chamber and/or ROXS reduces photobleaching, it is important to investigate if this modification is beneficial for gathering microarray data in general. To shed light on this subject, the spot intensity level as well as the spot's initial intensity deviation of unprotected spots were compared to their protected counterparts (intensity level: see Figure 5a,c,e, intensity deviation: see Supplementary Materials, Figure S1). The average intensity level seems to decrease for protected spots in some cases (Figure 5c,e) and seems to increase in others (Figure 5a,e). However, because of the overall broad distribution of intensity levels within each group, these decreases and increases are never significant, even for  $\alpha = 0.1$ . Comparing the intensity deviations of protected and unprotected spots by Analysis of Variance (ANOVA) showed no significant change.



**Figure 5.** Examination of the influence of protective measures on the mean spot intensity level and the spot intensity percent change after 10 scans. Intensity levels are derived from Array1: unprotected vs. PBS (a), Array2: unprotected vs. 10 mM ROXS (c), and Array3: unprotected vs. 50 mM ROXS (e). Spot intensity percent changes after 10 scans are derived from Array1: unprotected vs. 1 × PBS (b), Array2: unprotected vs. 10 mM ROXS (d), and Array3: unprotected vs. 50 mM ROXS (f). Intensity information derived from hybridizations of Cy3/Cy5 vs. unlabeled cDNA is tagged “single-dye” while intensity information derived from hybridization of Cy3-labeled cDNA with Cy5-labeled cDNA is tagged “as two-dye”. Error indicators are the respective standard errors of the mean (confidence: 95.4%) with Gaussian error propagation.

The evaluation of spot intensity percent change for unprotected vs. PBS shows that, for all but one combination of dyes, no significant change in intensity percent change is observable (for  $\alpha \leq 0.1$ ). Only for Cy5 vs. unlabeled the application of the liquid chamber significantly ( $\alpha = 0.05$ ) decreases the average intensity percent change from  $-6.51\%$  ( $\pm 5.47\%$ ) to  $0.18\%$  ( $\pm 1.45\%$ ). Conversely, for unprotected vs. 10 mM ROXS, all dye combinations see a significant ( $\alpha \leq 0.05$ ) elevation of percent change levels. Most of them are even significant for  $\alpha \leq 0.01$ . While the application of 50 mM ROXS leads to significant reduction of intensity loss for Cy3-labeled DNAs ( $\alpha \leq 0.01$ ), this cannot be concluded for their Cy5-labeled counterparts. It should be noted that percent changes for Cy5 on this array were lower in general, compared to the other arrays.

While there are singular significant differences when comparing the intensity level and intensity percent change of a labeled DNA of a single dye spot with its two-dye counterpart, the overall results show that no significant differences ( $\alpha \leq 0.1$ ) exist in this 24 gene experiment. Similar to the ROXS results, the statistical analysis of the intensity level is limited by the overall broad dynamic range of intensity level within each group.

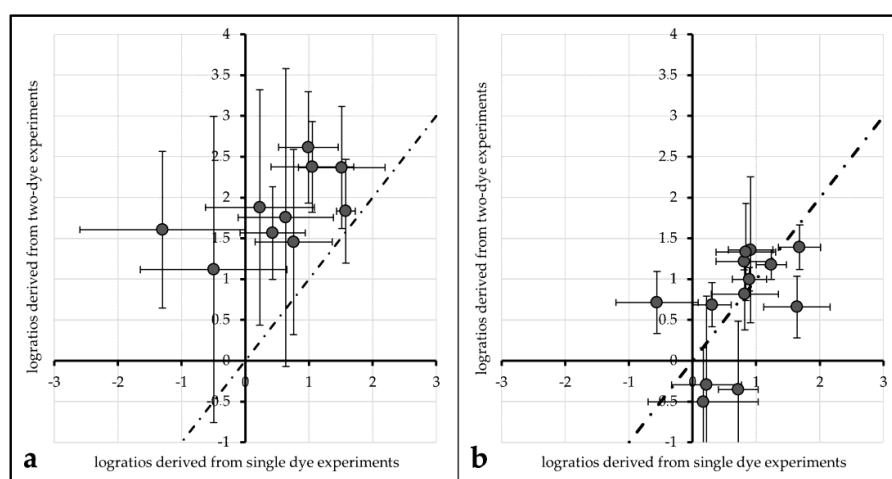
An evaluation of the impact of FRET on actual log-ratios was carried out to investigate the impact of FRET and/or protective measures on actual microarray data analysis. Log ratios (with base 2) of single labeled Cy3 and single labeled Cy5 spots were compared to log ratios derived from their respective Cy3. vs. Cy5 labeling counterparts.

$$\log_2 \text{ratio} (\text{Cy5}/\text{Cy3}) = \log_2 \left( \frac{\text{Intensity}_{\text{Cy5}}}{\text{Intensity}_{\text{Cy3}}} \right) \quad (2)$$

For unprotected spots of Array2, the average log-ratio derived from Cy3-single intensity divided by Cy5-single intensity was  $0.54$  ( $\pm 0.45$ ), while the average log-ratio from two-dye spots was  $1.86$  ( $\pm 0.66$ ). These two values differ significantly ( $\alpha \leq 0.01$ ). Plotted against each other, all data points lie above a line from the origin with a slope of 1 (see Figure 6a).

This tendency can also be observed for the same comparisons made with  $1 \times$  PBS protected spots of Array1 (single dye log ratio:  $0.75$  ( $\pm 0.28$ ), two-dye log ratio:  $1.53$  ( $\pm 0.23$ )) and unprotected spots of Array2 (single dye log ratio:  $0.72$  ( $\pm 0.35$ ), two-dye log ratio:  $1.64$  ( $\pm 0.37$ )). For both treatments, single dye log ratios are significantly different ( $\alpha \leq 0.01$ ) from two-dye log ratios.

Comparing the same values for ROXS-treated spots of Array2 gives different results: the mean log ratios of single dyes ( $0.99$  ( $\pm 0.23$ )) are not significantly different from those of two-dye spots ( $0.91$  ( $\pm 0.26$ )) for  $\alpha \leq 0.1$  (graphical representation: Figure 6b).



**Figure 6.** Comparison of log<sub>2</sub>-ratios derived from single-dye spots (635 nm intensity from Cy3 vs. unlabeled, 532 nm intensity from unlabeled vs. Cy5) plotted against log<sub>2</sub>-ratios derived from two-dye spots for unprotected spots of Array2 (a) and spots protected by 10 mM ROXS in PBS (b). The dotted line passes the origin with a slope of 1.

**Table 1.** Results and statistical metadata regarding the mean initial spot intensity (I0) of the 24 gene experiment.

Array	Dye Combination	Mean I0	Std. Error $\alpha = 95.4\%$	Mean I0	Std. Error $\alpha = 95.4\%$	p-Value
1		unprotected		PBS		
	Cy3 single dye	$6.08 \times 10^3$	$1.96 \times 10^3$	$9.58 \times 10^3$	$5.30 \times 10^3$	$2.28 \times 10^{-1}$
	Cy5 single dye	$3.47 \times 10^3$	$1.34 \times 10^3$	$6.37 \times 10^3$	$4.22 \times 10^3$	$2.03 \times 10^{-1}$
	Cy3 two dye	$7.14 \times 10^3$	$1.57 \times 10^3$	$1.36 \times 10^4$	$7.07 \times 10^3$	$8.79 \times 10^{-2}$
	Cy5 two dye	$1.74 \times 10^3$	$4.64 \times 10^2$	$5.39 \times 10^3$	$3.54 \times 10^3$	$5.26 \times 10^{-2}$
2		unprotected		10 mM ROXS in PBS		
	Cy3 single dye	$1.49 \times 10^4$	$9.73 \times 10^3$	$8.83 \times 10^3$	$5.98 \times 10^3$	$3.23 \times 10^{-1}$
	Cy5 single dye	$7.90 \times 10^3$	$4.24 \times 10^3$	$3.77 \times 10^3$	$1.68 \times 10^3$	$1.02 \times 10^{-1}$
	Cy3 two dye	$1.51 \times 10^4$	$1.00 \times 10^4$	$8.96 \times 10^3$	$4.74 \times 10^3$	$3.08 \times 10^{-1}$
	Cy5 two dye	$4.51 \times 10^3$	$2.98 \times 10^3$	$5.02 \times 10^3$	$2.77 \times 10^3$	$8.12 \times 10^{-1}$
3		unprotected		50 mM ROXS in PBS		
	Cy3 single dye	$9.48 \times 10^3$	$4.71 \times 10^3$	$1.47 \times 10^4$	$6.95 \times 10^3$	$2.58 \times 10^{-1}$
	Cy5 single dye	$1.18 \times 10^4$	$8.40 \times 10^3$	$6.63 \times 10^3$	$4.75 \times 10^3$	$3.73 \times 10^{-1}$
	Cy3 two dye	$7.39 \times 10^3$	$3.80 \times 10^3$	$1.55 \times 10^4$	$7.91 \times 10^3$	$9.03 \times 10^{-2}$
	Cy5 two dye	$6.46 \times 10^3$	$4.49 \times 10^3$	$7.97 \times 10^3$	$5.76 \times 10^3$	$7.08 \times 10^{-1}$

**Table 2.** Results and statistical metadata regarding the mean spot intensity percent change after 10 scans (relative  $\Delta I0$ ) of the 24 gene experiment.

Array	Dye Combination	Relative $\Delta I0$	Std. Error $\alpha = 95.4\%$	Relative $\Delta I0$	Std. Error $\alpha = 95.4\%$	p-Value
1		unprotected		PBS		
	Cy3 single dye	$-7.15 \times 10^{-2}$	$2.31 \times 10^{-2}$	$-6.88 \times 10^{-2}$	$1.37 \times 10^{-2}$	$8.45 \times 10^{-1}$
	Cy5 single dye	$-6.51 \times 10^{-2}$	$5.47 \times 10^{-2}$	$1.81 \times 10^{-2}$	$1.45 \times 10^{-2}$	$2.67 \times 10^{-2}$
	Cy3 two dye	$-3.68 \times 10^{-2}$	$2.97 \times 10^{-2}$	$-5.91 \times 10^{-2}$	$9.39 \times 10^{-3}$	$1.67 \times 10^{-1}$
	Cy5 two dye	$-4.41 \times 10^{-2}$	$4.28 \times 10^{-2}$	$-4.38 \times 10^{-2}$	$1.19 \times 10^{-2}$	$9.88 \times 10^{-1}$
2		unprotected		10 mM ROXS in PBS		
	Cy3 single dye	$-4.77 \times 10^{-2}$	$1.17 \times 10^{-2}$	$-6.77 \times 10^{-2}$	$9.30 \times 10^{-3}$	$1.68 \times 10^{-2}$
	Cy5 single dye	$-2.98 \times 10^{-2}$	$1.48 \times 10^{-2}$	$6.88 \times 10^{-3}$	$1.67 \times 10^{-2}$	$3.80 \times 10^{-3}$
	Cy3 two dye	$-5.94 \times 10^{-2}$	$1.19 \times 10^{-2}$	$-9.42 \times 10^{-3}$	$2.36 \times 10^{-2}$	$1.01 \times 10^{-3}$
	Cy5 two dye	$-3.16 \times 10^{-2}$	$1.54 \times 10^{-2}$	$-1.93 \times 10^{-3}$	$1.19 \times 10^{-2}$	$7.35 \times 10^{-3}$
3		unprotected		50 mM ROXS in PBS		
	Cy3 single dye	$-6.19 \times 10^{-2}$	$1.30 \times 10^{-2}$	$-3.60 \times 10^{-2}$	$8.31 \times 10^{-3}$	$8.59 \times 10^{-3}$
	Cy5 single dye	$-8.72 \times 10^{-3}$	$1.12 \times 10^{-2}$	$-2.84 \times 10^{-3}$	$9.58 \times 10^{-3}$	$4.89 \times 10^{-1}$
	Cy3 two dye	$-6.58 \times 10^{-2}$	$1.21 \times 10^{-2}$	$-3.10 \times 10^{-2}$	$9.46 \times 10^{-3}$	$6.78 \times 10^{-4}$
	Cy5 two dye	$-1.95 \times 10^{-2}$	$8.40 \times 10^{-3}$	$-1.38 \times 10^{-2}$	$6.60 \times 10^{-3}$	$3.64 \times 10^{-1}$

#### 4. Discussion

The importance of reliable bias normalization and quality control is well recognized in the microarray field. Next to biological and biochemical sources, bias originates from photochemical processes and depends on the choice of labeling agent as well as the selected imaging procedure and environment. In earlier works, it was shown that the ubiquitous application of cyanine dye labeling causes significant bias due to the dyes' disparate susceptibility to photobleaching and possible FRET interaction [19,20,26,28,29].

These findings are confirmed in this study, as photobleaching and FRET result in significantly different data: as shown in Figure 4(1a,b), without protective measures, photobleaching occurs similar to previous findings of von der Haar et al. [19] with intensity decreases for Cy3 and to a higher degree for Cy5. Interestingly, the 96 gene experiment shows that these photobleaching percentages nearly switch when applying a wet chamber with ROXS. A comparably higher decrease for Cy3 intensity and a lower intensity decrease for Cy5 is observed. Both changes are statistically significant. These findings

confirm those of Vogelsang et al. [20], especially the reduction of Cy5 intensity loss. The increase of Cy3-intensity loss is significantly stronger for Cy3-labeled DNA hybridized against Cy5-labeled DNA. It can be primarily traced back to a hypothesized FRET effect. This passive “de-quenching” effect, which has similarly been described by Rao et al. [26], was observable for a variety of genes, independent of the initial spot intensity (Supplementary Materials, SF1). In one case (Figure 4(1)), the higher photobleaching susceptibility of Cy5 most probably decreased the chance of excited Cy3 electrons to pass their energy over to nearby Cy5 molecules through FRET. This would result in a higher Cy3 emission that partially negates the intensity-decreasing effect of its own Cy3-photobleaching. In the presence of ROXS (Figure 4(2)), however, Cy5 would especially be protected from photo-destruction (see Rao et al. [26]). This would keep the rate of FRET between the two dyes stable, resulting in a visibly bigger and statistically significant intensity decrease of Cy3, which is not masked by “de-quenching” effects, as seen in Figure 4. Therefore, it is assumed that the de-quenching was observable not because of the selective bleaching of one cyanine dye.

Much of the reasoning applied above is based on the hypothesis that FRET happens to an observable degree in microarray experiments. In order to confirm FRET and quantify the effect, the experiment design was adapted so that for each bleaching condition there were spots with only Cy3-labeled DNA hybridized against unlabeled DNA, only Cy5-labeled vs. unlabeled, as well as Cy3-labeled vs. Cy5-labeled. In Figure 4(1a), the percentage of intensity change of the single dye spots after 10 scans is plotted against the same value for two-dye-spots for both cyanine dyes, respectively. If FRET does not occur to an observable degree, the absence/presence of a second dye would have no influence on the intensity change of the first. Data points for both dyes should then be scattered normally distributed around a spot/gene-specific point on a line from the origin with a slope of  $-1$ ; however, this is only the case for Cy3. It was found that unprotected Cy3-labeled DNA did not show a significantly different intensity decrease depending on the presence/absence of Cy5-labeled DNA. It is hypothesized that the overall low observable bleaching of this group’s spots obscures possible FRET effects. For unprotected Cy5-labeled DNA, however, the intensity decrease is significantly higher when Cy3-labeled DNA is present. A possible explanation is the higher rate of excitation of Cy5 due to FRET which subsequently leads to more chances of Cy5-photobleaching. For ROXS-protected spots, we see significantly higher bleaching for two-dye spots of both Cy3 and Cy5. While the explanation for a visible hypothesized FRET effect on Cy3 has been stated above, the question of why a hypothesized FRET effect is observable for Cy5 remains. Compared to unprotected Cy3 DNA, it is expected that the comparably overall low level of protected Cy5 intensity decrease results in a similar insignificant observable difference of intensity decrease of single-dye and two-dye spots. The fact that a significantly higher decrease is still observable for Cy5 of two-dye spots can be explained by referring to the cDNA labeling and scanner settings: although Cy5-molecules are only different in structure by one conjugated C-C double bond, Cy5’s direct labeling efficiency is significantly lower compared to the one of Cy3. This source of possible bias is mostly addressed by adjusting/increasing the photomultiplier voltage of the 635 nm laser to lift the intensity level of Cy5-signals to the one of Cy3-signals. As the process of photo multiplication exponentially enhances the photon signal, the comparably higher voltage applied to Cy5-emitted photons might lead to a non-linear enhancement of intensity resolution. Consequently, two differently intense Cy5 photon signals might result in a larger observed intensity difference than that of two equally different Cy3 signals. While a stronger bleaching of Cy5 is to be expected, the data of the 24 gene experiment does not support this hypothesis. In all cases, Cy3 loses a higher fraction of its intensity after ten scans. Concerning the setups with ROXS present, this protective buffer has a stronger preserving effect on Cy5 than on Cy3, as described in the literature [20]. This could result in comparably higher observable relative Cy3 intensity loss. The results of unprotected spots of the 24 gene experiment should show a higher relative intensity change for Cy5, as they do in the 96 gene experiment (Figure 5b). Why this is not the case remains unclear.

Overall, this experiment statistically supports the hypothesis that FRET is an observable effect in DNA microarrays. Statistical inquiries showed that the percentage intensity change evaluated here

is independent of the initial intensity level of the spot under study, ruling out a possible intensity level bias, as first evaluations showed that the overall intensity level is decreased by applying a liquid chamber onto the array (additional data can be found in the Supplementary Materials, Tables S2 to S4).

After the evaluation of this experiment, questions remained: If the increase of observable Cy3-intensity loss is explained by FRET effects, why does it also occur, to a lesser extent, in single-dye setups? To what extent is the observed reduction of photobleaching caused by ROXS? Are the dyes merely protected due to the application of the liquid chamber itself? This would indicate that the bleaching is mostly caused by environmental ozone that is now efficiently blocked. A first evaluation of the 24 gene experiment showed that the data derived from unprotected spots of all three tested arrays are statistically comparable with respect to intensity level, percentage intensity change, and overall spot intensity standard deviation. This, in theory, allows for further examination and comparison of spots from different arrays; therefore, additional array-to-array normalization was not carried out, which might also introduce bias obfuscating FRET and/or ROXS effects.

The application of a liquid chamber did not influence the overall intensity levels or the magnitude of the dynamic signal intensity range for both dyes. While this is a desirable outcome regarding the intensity level, an effect on the dynamic range would have been a mixed blessing: increasing the dynamic range benefits the resolution and therefore the distinguishability on the one hand, but increases the need for problematic multi-scan applications to cover this broadened range on the other. A decrease would have the contrary effect, sacrificing resolution for more convenient scanning.

In order to ensure that the observed photobleaching protection is due to the ROXS buffer components, the 24 gene experiment's setup included spot protected by a liquid chamber filled with PBS buffer without ROXS. All comparison of intensity percent change made for unprotected vs. PBS protected spots showed no significant reduction of intensity loss through application of a liquid chamber with PBS buffer. The only exception was the single Cy5 DNA, though for a less significant threshold of  $\alpha = 0.05$ . Comparing unprotected spots with spots protected by a liquid chamber filled with 10 mM ROXS in PBS, however, displayed a highly significant reduction of intensity loss for all compared configurations. These findings strongly indicate that the presence of 10 mM ROXS is actually responsible for the changes observed in the 96 gene experiment. The additional test of unprotected vs. 50 mM ROXS did not yield conclusive results as a significant reduction of intensity loss was only observed for Cy3 and not for Cy5, though overall low intensity of Array3's spots might have affected the statistical power of these specific results. On the other hand, a 10 mM solution of ROXS is closer to the 1 mM formula used by Vogelsang et al. [20].

In contradiction with the 96 gene experiment, the application of ROXS in the 24 gene experiment also significantly reduced bleaching of Cy3-labeled DNA. Whether this change in observed behavior was due to the changed ROXS concentration or merely resulted from the absence of bias due to the better statistical power of the 96 gene experiment's design cannot be ascertained at this point. All in all, these results show that the application of a liquid chamber filled with a 1 mM or 10 mM ROXS solution provides a practical solution for significant reduction of cyanine dye photobleaching caused by DNA microarray scanning.

Regarding FRET, the same parameters used to evaluate this effect in the 96 gene experiment do not yield the expected results in the 24 gene experiment. Only 2 out of 12 comparisons showed a significantly different intensity percent change of a dye depending on the presence/absence of its cyanine counterpart. This might mislead the observer to the conclusion that the FRET influence observed in the 96 gene experiment is a bias which disappeared due to the better statistical power. A closer investigation of the effect FRET has on the results of a typical analysis carried out with the 24 gene experiment's data gives a different picture:  $\log_2$ -ratios derived from single-dye data in comparison with two-dye data were plotted against each other (Figure 6). For data derived from unprotected and PBS protected spots, the data points do not seem to be normally distributed around a line to the origin with a slope of 1. Normal distribution around this slope would be the expected result if the spots were not affected by FRET. This impression is statistically proofed as the means of

single-dye spots of unprotected/PBS-protected spots are significantly different from those of two-dye spots of the same treatment. Carrying out the same comparison for ROXS-protected spots of Array2 gives a different result: the mean  $\log_2$ -ratios of single-dye spots are not significantly different from their two-dye equivalents. These observations support the theory that FRET is not only occurring in two-dye microarrays, it is significantly biasing the results of these experiments. Furthermore, the FRET-induced bias seems to be normalized by applying the ROXS-protection, as no significant difference can be observed in this case. Therefore, the application of a ROXS-filled liquid chamber seems not only to be beneficial in terms of photobleaching minimization but also poses a valid strategy in order to normalize FRET-dependent bias in two-dye experiments.

In order to allow for this novel technology to be used in daily experiments, several investigations and optimizations remain. Is the remaining variability of  $\log_2$ -ratios caused by the difference of the treatments/dye usages or by systematic/technical variance? If the application of ROXS does actually compensate the bias introduced by FRET, as implicated by the 24 gene experiment's results, further investigations are necessary. Does ROXS minimize the occurrence of FRET by minimizing the availability of the specific excited electron state from which FRET is initiated in Cy3, much like with the photobleaching initiating states? Or, is FRET still occurring but the presence of ROXS implements another compensating effect? FRET is described as induced oscillation of two excited singlet-state electrons, while photodestruction originates from a triplet-state. Further replication of the experiments carried out in this study is needed to allow for quantification of FRET influences, leading to predictive models. Additionally, further tests are necessary to determine the optimal ROXS and buffer concentrations for DNA experiments. Future research should also broaden the application of this approach to protein and cellular microarrays. This necessitates the examination of ROXS's effect on protein-stability and the compounds' biocompatibility.

## 5. Conclusions

Based on the findings of Vogelsang et al. [20], Rao et al. [26], and our own previous research [19], a novel strategy for the minimization of photobleaching in cyanine-labeling-based DNA microarray experiments was successfully implemented. The modification of DNA microarray slides with thin liquid chambers filled with a buffer containing ROXS provided a valid protection of cyanine dyes against photo destruction occurring in the scanning process. Furthermore, it was shown that while FRET does not only occur in DNA microarray experiments, it does significantly bias the results of two-dye microarray derived data. This bias can successfully be normalized by applying the same ROXS-buffer-filled liquid chamber to the microarray slide. With necessary further optimization of this technology, the photonic limitations of cyanine-based microarray scanning can be overcome. This does not only improve the reproducibility of these experiments, it allows for successful implementation of multi-scan approaches with all the resulting possibilities.

**Supplementary Materials:** The following are available online at [www.mdpi.com/2079-7737/5/4/47/s1](http://www.mdpi.com/2079-7737/5/4/47/s1), Figure S1: Influence of spot intensity level on spot intensity percent change after 10 scans. (a) unprotected single dye spots; (b) unprotected two dye spots; (c) 1 mM ROXS in 1 mM PBS protected single dye spots; (d) 1 mM ROXS in 1 mM PBS protected two dye spots. Error indicators are simple standard deviations; Table S1: Sets of Genes for which oligos were designed as used in the 96 gene, two-array experiment and the 24 gene, three-array experiment; Table S1: Influence of presence absence of protective measures on overall spot intensity deviations for Array1 (unprotected vs. 1 mM PBS). *SS*: Sum of Squares, *df*: degrees of freedom, *MS*: Mean of Square Sums, *F*: F-value, *p*: p-value corresponding to *F*, *F<sub>crit</sub>*: critical *F* corresponding to chosen confidence interval ( $\alpha = 0.05$ ); Table S2: Influence of presence absence of protective measures on overall spot intensity deviations for Array2 (unprotected vs. 10 mM ROXS in 1 mM PBS). *SS*: Sum of Squares, *df*: degrees of freedom, *MS*: Mean of Square Sums, *F*: F-value, *p*: p-value corresponding to *F*, *F<sub>crit</sub>*: critical *F* corresponding to chosen confidence interval ( $\alpha = 0.05$ ); Table S3: Influence of presence absence of protective measures on overall spot intensity deviations for Array2 (unprotected vs. 10 mM ROXS in 1 mM PBS). *SS*: Sum of Squares, *df*: degrees of freedom, *MS*: Mean of Square Sums, *F*: F-value, *p*: p-value corresponding to *F*, *F<sub>crit</sub>*: critical *F* corresponding to chosen confidence interval ( $\alpha = 0.05$ ); Table S4: Influence of presence absence of protective measures on overall spot intensity deviations for Array3 (unprotected vs. 50 mM ROXS in 1 mM PBS).



**Author Contributions:** Marcel von der Haar, Martin Pähler, Patrick Lindner, and Frank Stahl conceived and designed the experiments. Marcel von der Haar and Christopher Heuer performed the experiments. Marcel von der Haar, Kathrin von der Haar, Patrick Lindner, Thomas Scheper, and Frank Stahl analyzed the data. Marcel von der Haar wrote the manuscript. All authors participated in the design of the study and approved the final manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Spielbauer, B.; Stahl, F. Impact of Microarray Technology in Nutrition and Food Research. *Mol. Nutr. Food Res.* **2005**, *49*, 908–917. [[CrossRef](#)] [[PubMed](#)]
2. Allison, D.B.; Cui, X.; Page, G.P.; Sabripour, M. Microarray Data Analysis: From Disarray to Consolidation and Consensus. *Nat. Rev. Genet.* **2006**, *7*, 55–65. [[CrossRef](#)] [[PubMed](#)]
3. Ehrenreich, A. DNA Microarray Technology for the Microbiologist: An Overview. *Appl. Microbiol. Biotechnol.* **2006**, *73*, 255–273. [[CrossRef](#)] [[PubMed](#)]
4. Kretschy, N.; Somoza, M.M. Comparison of the Sequence-Dependent Fluorescence of the Cyanine Dyes Cy3, Cy5, Dylight Dy547 and Dylight Dy647 on Single-Stranded DNA. *PLoS ONE* **2014**, *9*, e85605. [[CrossRef](#)] [[PubMed](#)]
5. Mary-Huard, T.; Daudin, J.J.; Robin, S.; Bitton, F.; Cabannes, E.; Hilson, P. Spotting Effect in Microarray Experiments. *BMC Bioinform.* **2004**. [[CrossRef](#)] [[PubMed](#)]
6. Dawson, E.D.; Reppert, A.E.; Rowlen, K.L.; Kuck, L.R. Spotting Optimization for Oligo Microarrays on Aldehyde-Glass. *Anal. Biochem.* **2005**, *341*, 352–360. [[CrossRef](#)] [[PubMed](#)]
7. Sobek, J.; Aquino, C.; Weigel, W.; Schlapbach, R. Drop Drying on Surfaces Determines Chemical Reactivity—The Specific Case of Immobilization of Oligonucleotides on Microarrays. *BMC Biophys.* **2013**. [[CrossRef](#)] [[PubMed](#)]
8. Rao, A.N.; Grainger, D.W. Biophysical Properties of Nucleic Acids at Surfaces Relevant to Microarray Performance. *Biomater. Sci.* **2014**, *2*, 436–471. [[CrossRef](#)] [[PubMed](#)]
9. Jang, H.; Cho, M.; Kim, H.; Kim, C.; Park, H. Quality Control Probes for Spot-Uniformity and Quantitative Analysis of Oligonucleotide Array. *J. Microbiol. Biotechnol.* **2009**, *19*, 658–665. [[PubMed](#)]
10. Khondoker, M.R.; Glasbey, C.A.; Worton, B.J. Statistical Estimation of Gene Expression Using Multiple Laser Scans of Microarrays. *Bioinformatics* **2006**, *22*, 215–219. [[CrossRef](#)] [[PubMed](#)]
11. Satterfield, M.B.; Lippa, K.; Lu, Z.Q.; Salit, M.L. Microarray Scanner Performance over a Five-Week Period as Measured with Cy5 and Cy3 Serial Dilution Slides. *J. Res. Natl. Inst. Stand. Technol.* **2008**, *113*, 157–174. [[CrossRef](#)] [[PubMed](#)]
12. Ambroise, J.; Bearzatto, B.; Robert, A.; Macq, B.; Gala, J.L. Combining Multiple Laser Scans of Spotted Microarrays by Means of a Two-Way Anova Model. *Stat. Appl. Genet. Mol. Biol.* **2012**. [[CrossRef](#)] [[PubMed](#)]
13. Vora, G.J.; Meador, C.E.; Anderson, G.P.; Taitt, C.R. Comparison of Detection and Signal Amplification Methods for DNA Microarrays. *Mol. Cell. Probes* **2008**, *22*, 294–300. [[CrossRef](#)] [[PubMed](#)]
14. Shi, L.; Tong, W.; Su, Z.; Han, T.; Han, J.; Puri, R.K.; Fang, H.; Frueh, F.W.; Goodsaid, F.M.; Guo, L.; et al. Microarray Scanner Calibration Curves: Characteristics and Implications. *BMC Bioinform.* **2005**. [[CrossRef](#)] [[PubMed](#)]
15. Lyng, H.; Badiie, A.; Svendsrud, D.H.; Hovig, E.; Myklebost, O.; Stokke, T. Profound Influence of Microarray Scanner Characteristics on Gene Expression Ratios: Analysis and Procedure for Correction. *BMC Genom.* **2004**. [[CrossRef](#)]
16. Dar, M.; Giesler, T.; Richardson, R.; Cai, C.; Cooper, M.; Lavasani, S.; Kille, P.; Voet, T.; Vermeesch, J. Development of a Novel Ozone- and Photo-Stable Hyper5 Red Fluorescent Dye for Array Cgh and Microarray Gene Expression Analysis with Consistent Performance Irrespective of Environmental Conditions. *BMC Biotechnol.* **2008**. [[CrossRef](#)] [[PubMed](#)]
17. Kuang, C.; Luo, D.; Liu, X.; Wang, G. Study on Factors Enhancing Photobleaching Effect of Fluorescent Dye. *Measurement* **2013**, *46*, 1393–1398. [[CrossRef](#)]
18. Drăghici, S. *Statistics and Data Analysis for Microarrays Using R and Bioconductor*, 2nd ed.; Taylor & Francis: Abingdon, UK, 2011.
19. Von der Haar, M.; Preuß, J.A.; von der Haar, K.; Lindner, P.; Scheper, T.; Stahl, F. The Impact of Photobleaching on Microarray Analysis. *Biology* **2015**, *4*, 556–572. [[CrossRef](#)] [[PubMed](#)]

20. Vogelsang, J.; Kasper, R.; Steinhauer, C.; Person, B.; Heilemann, M.; Sauer, M.; Tinnefeld, P. A Reducing and Oxidizing System Minimizes Photobleaching and Blinking of Fluorescent Dyes. *Angew. Chem. Int. Ed. Engl.* **2008**, *47*, 5465–5469. [[CrossRef](#)] [[PubMed](#)]
21. Widengren, J.; Chmyrov, A.; Eggeling, C.; Lofdahl, P.A.; Seidel, C.A.M. Strategies to Improve Photostabilities in Ultrasensitive Fluorescence Spectroscopy. *J. Phys. Chem. A* **2007**, *111*, 429–440. [[CrossRef](#)] [[PubMed](#)]
22. Van der Velde, J.H.; Oelerich, J.; Huang, J.; Smit, J.H.; Hiermaier, M.; Ploetz, E.; Herrmann, A.; Roelfes, G.; Cordes, T. The Power of Two: Covalent Coupling of Photostabilizers for Fluorescence Applications. *J. Phys. Chem. Lett.* **2014**, *5*, 3792–3798. [[CrossRef](#)] [[PubMed](#)]
23. Sapsford, K.E.; Berti, L.; Medintz, I.L. Materials for Fluorescence Resonance Energy Transfer Analysis: Beyond Traditional Donor-Acceptor Combinations. *Angew. Chem. Int. Ed.* **2006**, *45*, 4562–4588. [[CrossRef](#)] [[PubMed](#)]
24. Sabanayagam, C.R.; Eid, J.S.; Meller, A. Using fluorescence resonance energy transfer to measure distances along individual dna molecules: Corrections due to nonideal transfer. *J. Chem. Phys.* **2005**. [[CrossRef](#)] [[PubMed](#)]
25. Dinant, C.; van Royen, M.E.; Vermeulen, W.; Houtsmuller, A.B. Fluorescence Resonance Energy Transfer of Gfp and Yfp by Spectral Imaging and Quantitative Acceptor Photobleaching. *J. Microsc.* **2008**, *231*, 97–104. [[CrossRef](#)] [[PubMed](#)]
26. Rao, A.N.; Rodesch, C.K.; Grainger, D.W. Real-Time Fluorescent Image Analysis of DNA Spot Hybridization Kinetics to Assess Microarray Spot Heterogeneity. *Anal. Chem.* **2012**, *84*, 9379–9387. [[CrossRef](#)] [[PubMed](#)]
27. Qin, L.X.; Kerr, K.F.; Contributing Members of the Toxicogenomics Research Consortium. Empirical Evaluation of Data Transformations and Ranking Statistics for Microarray Analysis. *Nucleic Acids Res.* **2004**, *32*, 5471–5479. [[CrossRef](#)] [[PubMed](#)]
28. Rao, A.N.; Rodesch, C.K.; Grainger, D.W. Supporting Information of Real-Time Fluorescent Image Analysis of DNA Spot Hybridization Kinetics to Assess Microarray Spot Heterogeneity (Vol 84, Pg 9379, 2012). *Anal. Chem.* **2013**, *85*, 4199. [[CrossRef](#)]
29. Rao, A.N.; Vandencastele, N.; Gamble, L.J.; Grainger, D.W. High-Resolution Epifluorescence and Time-of-Flight Secondary Ion Mass Spectrometry Chemical Imaging Comparisons of Single DNA Microarray Spots. *Anal. Chem.* **2012**, *84*, 10628–10636. [[CrossRef](#)] [[PubMed](#)]



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).

### 3.3 Array Analysis Manager – An automated DNA microarray analysis tool simplifying microarray data filtering, bias recognition, normalization and expression analysis

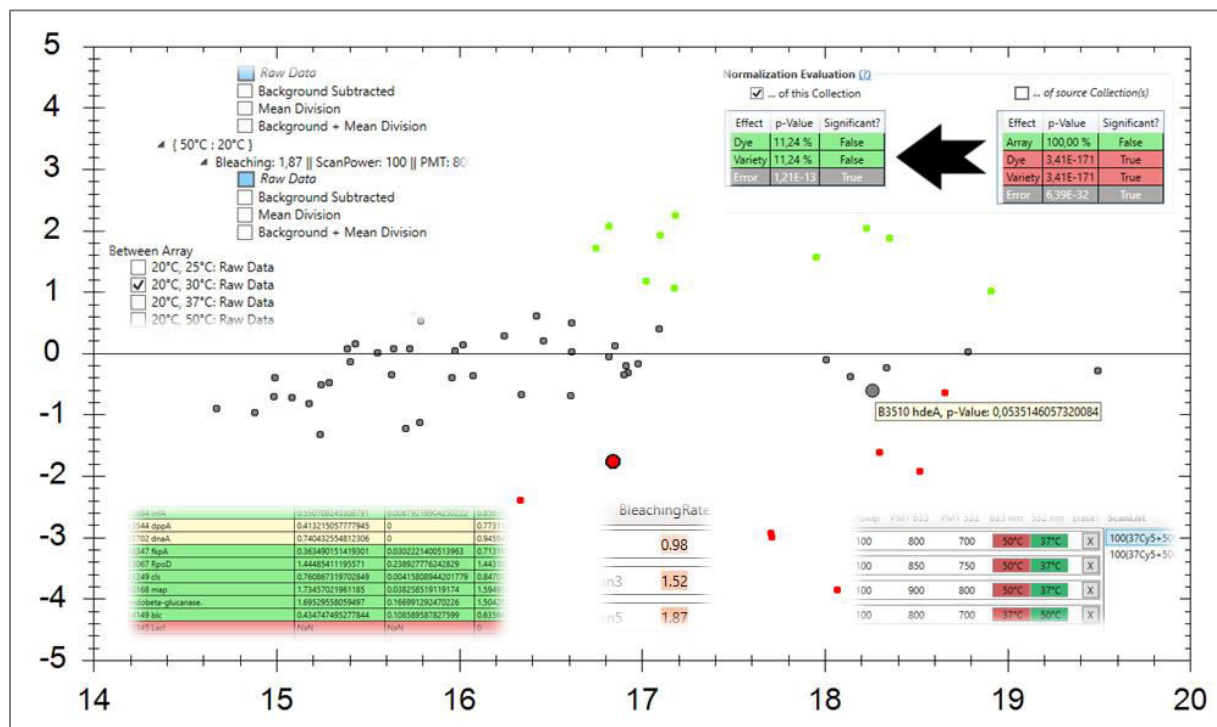


Figure 3.3: Graphical abstract of "Array Analysis Manager – An automated DNA microarray analysis tool simplifying microarray data filtering, bias recognition, normalization and expression analysis"

This section describes the *Array Analysis Manager* end user software. The program is designed to contribute to reproducibility and comparability of microarray experiment results. It is compatible with basic single chip experimental design as well as dye swap and loop approaches. Additionally, multi-scan datasets can be processed to harness the potential of enhanced dynamic intensity range evaluations. The *Array Analysis Manager* is the first software equipped with photobleaching recognition and normalization algorithms [68-71]. The normalization algorithm is an implementation of the empirical model explained in chapter 3.1. The recognition function was developed using an ANN, provided with bleaching data gathered from low density two-channel microarrays. Also, an ANOVA-based normalization evaluation tool was designed and implemented. It offers vital feedback on how efficient several bias sources are normalized by all supported normalization techniques and their combinations.

ANOVA is used to evaluate the composition of a regression's variance in order to determine if and how much variance is introduced by independent variables/factors. It customarily works with sums of squares. The initial total sum of squares  $SS_{\text{Total}}$  is made up from the sum of squares of the mean  $SS_{\text{Mean}}$  and the sum corrected by the mean  $SS_{\text{Corrected}}$ . The latter can be divided into the sum

## Experimental Part

of squares of the effects of independent variables / factors  $SS_{\text{Factors}}$  and the residual sum of squares  $SS_{\text{R}}$ . The residual sum can be further divided into the model deviation  $SS_{\text{Model}}$  and the experimental error  $SS_{\text{Exp.-Error}}$  [72]. Calculations for these sums of squares are given in Table 3.1. This categorization of the variances of a dataset can be used to gain a variety of indicators. When used for regression analysis quantities such as the coefficients of determination and correlation can be determined. It further allows for goodness-of-fit and lack-of-fit testing as well as the determination of confidence intervals for the factors under investigation. What makes it interesting in terms of normalization evaluation is the fact that it allows to separate the deviation derived from independent factors from the residual error. This, in theory allows for a comparison of these two deviations analogue to an F-test. Given that the residual error is normally distributed, a significant difference of these two deviations can be an indicator that the factor under study affects the dataset [39].

**Table 3.1:** Calculations of sums of squares (SS) for a linear regression's analysis of variance (ANOVA).  $y$  - dependent variable;  $j$  - vector containing all  $y$ -means of the  $i^{\text{th}}$  measurement/observation;  $f$  - number of factor combinations;  $n$  - number of measurements/observations;  $p$  - number of independent parameters;  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$  - total mean [72]

sum of squares	matrix operation	calculation	degrees of freedom
$SS_{\text{Total}}$ , total	$y^T \times y$	$\sum_{i=1}^n y_i^2$	$n$
$SS_{\text{Mean}}$ , mean	$\bar{y}^T \times \bar{y}$	$n \times \bar{y}^2$	$1$
$SS_{\text{Corrected}}$ , corrected by mean	$(y - \bar{y})^T (y - \bar{y})$	$\sum_{i=1}^n (y_i - \bar{y})^2$	$n - 1$
$SS_{\text{Factors}}$ , factors	$(\hat{y} - \bar{y})^T (\hat{y} - \bar{y})$	$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	$p - 1$
$SS_{\text{R}}$ , residuals	$(y - \hat{y})^T (y - \hat{y})$	$\sum_{i=1}^n (y_i - \hat{y}_i)^2$	$n - p$
$SS_{\text{Model}}$ , model deviation	$(j - \hat{y})^T (j - \hat{y})$	$\sum_{i=1}^n (\bar{y}_i - \hat{y}_i)^2$	$f - p$
$SS_{\text{Exp.-Error}}$ , experimental error	$(y - j)^T (y - j)$	$\sum_{i=1}^n (y_i - \bar{y}_i)^2$	$n - f$

Analysis results can be exported in *.xlsx*- and *GEO*-formats, promoting easy and fast data processing and distribution. The overall user experience was designed to guide the experimenter through the analysis process. Tool tips provide information and guidelines for each step from data preprocessing, choice of normalization up to the detection of differentially expressed genes. This software was developed to increase the comparability and significance of microarray experiment result data by simplifying and harmonizing the data analysis process as well as providing experimenters of varying scientific and mathematic backgrounds with the information needed to make educated decisions about the processing of his experimental data.

Marcel von der Haar 

Patrick Lindner

Thomas Scheper

Frank Stahl

Institut für Technische Chemie,  
Leibniz Universität Hannover,  
Hannover, Germany

## Technical Report

# Array Analysis Manager—An automated DNA microarray analysis tool simplifying microarray data filtering, bias recognition, normalization, and expression analysis

Desoxyribonucleic acid (DNA) microarray experiments generate big datasets. To successfully harness the potential information within, multiple filtering, normalization, and analysis methods need to be applied. An in-depth knowledge of underlying physical, chemical, and statistical processes is crucial to the success of this analysis. However, due to the interdisciplinarity of DNA microarray applications and experimenter backgrounds, the published analyses differ greatly, for example, in methodology. This severely limits the comprehensibility and comparability among studies and research fields. In this work, we present a novel end-user software, developed to automatically filter, normalize, and analyze two-channel microarray experiment data. It enables the user to analyze single chip, dye-swap, and loop experiments with an extended dynamic intensity range using a multiscan approach. Furthermore, to our knowledge, this is the first analysis software solution, that can account for photobleaching, automatically detected by an artificial neural network. The user gets feedback on the effectiveness of each applied normalization regarding bias minimization. Standardized methods for expression analysis are included as well as the possibility to export the results in the Gene Expression Omnibus (GEO) format. This software was designed to simplify the microarray analysis process and help the experimenter to make educated decisions about the analysis process to contribute to reproducibility and comparability.

**Keywords:** ANOVA / Artificial Neural Networks / DNA microarrays / Photobleaching / Transcriptomics

*Received:* February 28, 2017; *revised:* April 18, 2017; *accepted:* May 16, 2017

**DOI:** 10.1002/elsc.201700046

## 1 Introduction

Microarray technology allows for efficient high-throughput transcriptome analysis, based on the competitive hybridization of complementary desoxyribonucleic acid (cDNA) probes of different regulatory states labeled with different fluorophores. The intensity ratios of these fluorophores correlate with the abundance of the original messenger ribonucleic acids and allow the experimenter to study a possible regulatory change of the genes under study [1]. This principle is widely used in fields such as medicine and molecular biology [2, 3].

**Correspondence:** Marcel von der Haar (koch@iftc.uni-hannover.de), Institut für Technische Chemie, Leibniz Universität Hannover, Callinstr. 5, 30167 Hannover, Germany

**Abbreviations:** ANOVA, analysis of variance; Cy3/5, cyanine 3/5; DNA, desoxyribonucleic acid; GEO, Gene Expression Omnibus; **gpr**, GenePix Pro results document; **.xlsx**, Microsoft Excel document

Next to technical and biochemical challenges, for example, regarding the choice of labeling agent or experimental setup the overall comparability of DNA microarray results is diminished by the heterogeneity of applied data analysis strategies. While steps have been taken to develop and establish common minimal analysis quality criteria for experimental design and data analysis [4–10], publications involved with DNA microarray technology still apply different analysis approaches and standards. Given the various backgrounds of the experimenters this is no surprise, as each field of research developed its own specialized set of evaluation procedures. Their statistical tools do often differ in nomenclature but also in preferred solution strategy. Several strategies were presented to break down these limitations. Pure mathematical/statistical algorithms correct saturation and/or noise using only the information lying in the acquired data itself. A Bayesian hierarchical model that corrects signal saturation based on pixel intensities was proposed by Gupta et al. [11]. Another strategy to overcome these limitations extends the scanning routine by

recording multiple scans while varying laser power and/or photomultiplier settings [12]. Khondoker et al. [13] use a maximum likelihood estimations model based on a Cauchy distribution to account for saturated signals and systematic bias. A scanner specific, photomultiplier-independent optical scanner bias was determined by Ambroise et al. [14]. They then constructed a two-way analysis of variance (ANOVA) model, normalizing scanner bias as well as saturation and noise. These techniques, among others, were reported to have increased the overall data quality and reproducibility compared to single scan designs [15, 16].

The need for a common and comprehensive analysis strategy has led to the development of multiple analysis tools. The *BAR-BAR* package for R programmers [17] enhances this programming language to normalize and analyze expression datasets. With *GEDI*, another approach is used. This software features a graphical user interface and several regression models used for normalization, expression analysis, clustering, classification, and construction of regulatory networks [18]. *quantro* is a software that evaluates the benefit of applying global normalization methods onto the dataset under study [19]. Next to analysis tools, solutions are developed for storage and management of microarray datasets. While the National Center for Biology Information (NCBI) solution Gene Expression Omnibus (*GEO*) [20] is widely used, alternative infrastructures such as *SaDA* emerge [21].

In this work, we present a novel software designed to establish a standardized set of solution strategies for DNA microarray analysis. It uses Microsoft's .NET-framework technology, allowing the software to be executed on any computer system with preinstalled .NET-framework. Working with GenePix Pro results document (.gpr) files, a widespread scanning result data format, our software assists the experimenter and facilitates the whole process of data filtering, normalization, and expression analysis all the way up to exporting the experimental results in the *GEO* format. Single array, single scan experiments can be analyzed as well as dye-swap and loop setups. Furthermore, the software supports experimental setups that use multiscan techniques to enhance the dynamic intensity range. The program includes an artificial neural network trained to recognize possible photobleaching and will recommend the application of the, to our best knowledge, very first correction algorithm, which is also included. Instead of simply applying a standard set of normalizations, the program uses an algorithm based on variance analysis to examine the effect of different normalization methods on possible bias sources. This allows the experimenter to make an informed decision on which normalizations should be applied. Regarding expression analysis, a set of standardized hypotheses tests enhanced by corrections for high test numbers automatically evaluates the relative expression using widely used statistical thresholds. This tool shall contribute to increase the overall comparability and reproducibility of DNA microarray experiment results to harness the full potential of this technology.

## 2 Implementation

The current version of the *Array Analysis Manager* (1.0) is freely available [22] runs on Windows operating systems and requires the preinstallation of the .NET-Framework environment which is freely available [23]. The *Array Analysis Manager* was developed

entirely in C# using the .NET development environment Visual Studio 2012 (Microsoft Corporation, Redmond, WA, USA) which is also freely available [24].

## 3 Key software features

### 3.1 Multiscan support

All single scan approaches are limited by the small dynamic intensity range of the available array scanners. The dynamic fluorescence intensity range of labeling agents, for example, cyanine dyes, goes well beyond what a single array scan can cover [25]. In most cases, feature information is lost due to saturation effects and noise at both ends of the intensity spectrum. Several strategies were presented to break down these limitations.

The *Array Analysis Manager* applies an approach suggested by Repenning [26]. If scans with at least three different photomultiplier voltage settings for at least one laser power setting are given for each chip of the experiment under study, the software automatically uses its multiscan algorithm. The algorithm applies an extended gamma correction to exclude data points from outside the linear range. Given that a sufficient number of data points per feature remain, a linear regression is carried out for each feature individually. With this regression a simulated intensity is calculated for each feature for a fixed photomultiplier voltage of 700 volt (V).

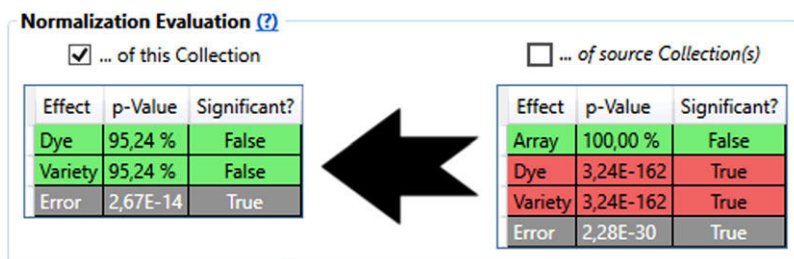
### 3.2 Photobleaching recognition and normalization

Photobleaching is an irreversible photochemical reaction destructing the fluorophores ability to emit photons [27]. The degree of signal loss differs depending on which fluorophore is used as a labeling agent [28–30]. Cyanine-3 (Cy3) and cyanine-5 (Cy5) are ubiquitously used for nucleotide labeling. In previous works, we showed that they possess a varying degree of susceptibility toward photobleaching [31]. All these findings imply an effect of photobleaching that will be even more significant on multiscan data quality. In the same previous work, we defined an empirical model suited for normalization of photobleaching introduced bias, given that the scanner settings and the order of applied scans are known (see Eq. (1), an in-detail discussion of this model can be found in von der Haar et al. [31]). The *Array Analysis Manager* is the first software equipped with a normalization algorithm based on this model that allows for automated photobleaching bias correction.

$$\ln(I(I_0, n_{\text{scan}}, V_{\text{PMT}})) = \ln(I_0) * e^{(-\lambda(V_{\text{PMT}}) * (n_{\text{scan}} - 1)^a(V_{\text{PMT}}))} \quad (1)$$

where  $I(I_0, n_{\text{scan}}, V_{\text{PMT}})$  is post PMT foreground intensity after  $n$  scans, with given  $I_0$  and  $V_{\text{PMT}}$ ;  $I_0$  is initial post PMT foreground intensity ( $n_{\text{scan}} = 1$ );  $n_{\text{scan}}$  is number of scans;  $V_{\text{PMT}}$  is PMT voltage;  $\lambda(V_{\text{PMT}})$  is degradation coefficient, and  $a(V_{\text{PMT}})$  is exponent.

In addition to normalizing photobleaching bias and evaluating the effect of this normalization (see Section 3.3), this software is equipped with a bleaching bias recognition function. This function is based on an artificial neural network trained



**Figure 1.** Exemplary screenshot of the result display of a comparative, ANOVA-based normalization evaluation. On the left, dye and variety bias are shown to be not significant, while the right side shows that the feature collection from which this collection was derived did show significant dye and variety bias but no significant array bias. From these results, the experimenter can abstract that the last applied normalization did successfully minimize dye and variety bias and that the choice of source collections was reasonable as no array bias was introduced by it. The informative value of these results, however, is diminished in this case, as the residual variance (systematic error) for both tested sets of feature collections did not pass the test for normal distribution. Screenshot taken on a Windows 10 system.

to recognize the relative degree of bleaching which a scan has already been exposed to. It was trained using metadata derived from low density two-channel scans with Cy3- and Cy5-labeled complementary DNA. Support for whole genome arrays and more labeling agents is planned for future updates. After .gpr-files are loaded into the program, each array will be evaluated by the neural network and relative bleaching grade numbers are assigned. While a coefficient of 1.0 equals an array that has not been scanned before, a coefficient of 2.0 indicates a Cy3 intensity loss of around 10% and/or a Cy5 loss of around 45% which is to be expected after five scans with 100% scan power. A coefficient of 3 and more represents intensity loss rate of around 20% (Cy3) and 80% (Cy5) and 10 or more previous scans. Given that scans are loaded into the program whose coefficients differ significantly, the program suggests the application of the bleaching bias normalization model.

### 3.3 ANOVA-based normalization evaluation

Many array analysis tools provide the user with multiple normalization algorithms. The decision, which algorithm is to be used, however, has to be made by the user. The lack of well-established standards, partially caused by the need for an individual approach for the respective experiment, leads to many parallel data handling methods. This mix complicates the transparency and reproducibility and therefore the scientific discourse and overall applicability of the microarray technology.

The *Array Analysis Manager* is equipped with multiple normalization algorithms, but also with a normalization evaluation function. This function is based on two-way ANOVA. The normalized dataset is analyzed before and after each normalization. Depending on the normalization and the amount of available data, the variance analysis compares the variance introduced by factors such as the labeling agents/dyes, the regulatory state, and/or the arrays with the residual variance which consists of the model deviation and the experimental error [32]. Given this residual variance is normally distributed, which too is tested using the Shapiro-Wilk test, a significant difference to this residual variance implies that the compared effect is significantly biasing the dataset. By comparing the  $p$ -values of the data before and after normalization, the experimenter gets information about

the influence of the applied normalization on these bias sources. After each applied normalization, the algorithm automatically calculates the aforementioned  $p$ -values and displays the results (Fig. 1). This provides the experimenter with useful feedback and arguments to choose the right normalization and explain his choice of data normalization methods.

### 3.4 Simplified workflow

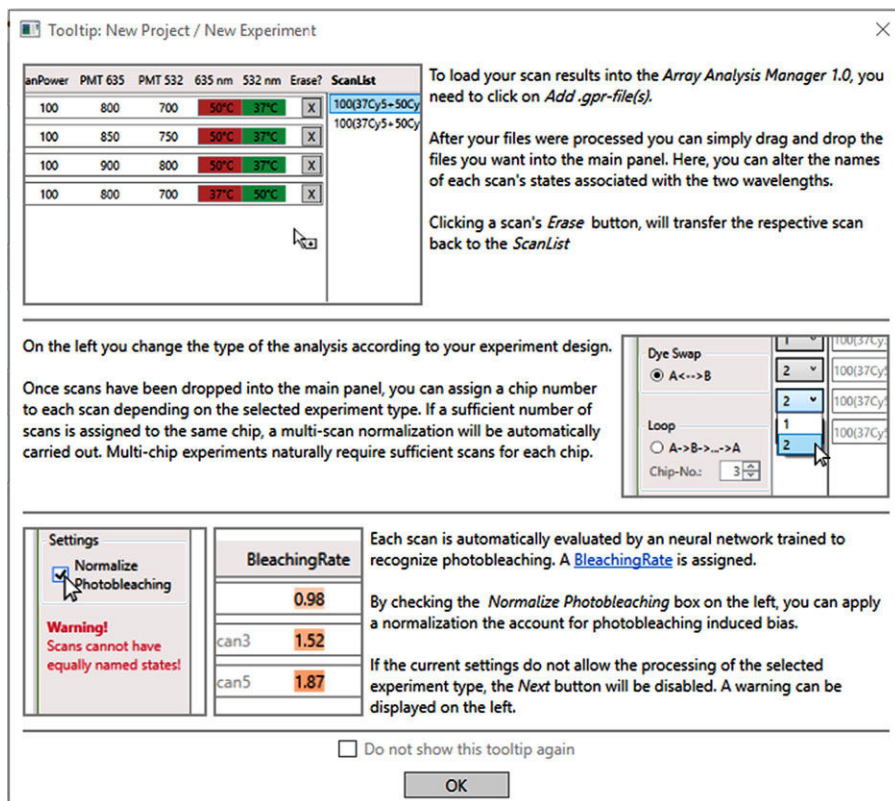
The *Array Analysis Manager* was designed with the objective of creating an easy-to-use, workflow optimizing assistant that gives the experimenter the tools and information needed to gain results quickly while making educated decisions regarding the processing of the array data.

The user experience should be intuitive and feasible, which is why instead of developing a toolbox for a programming language, it was decided to create a software with its own graphical user interface complete with drag and drop data handling. The application was created using *Microsoft's .NET environment* and the graphical user interface is based on the *Windows Presentation Foundation*, which will easily allow the implementation of a browser-based web version of the *Array Analysis Manager* for world-wide instant accessibility.

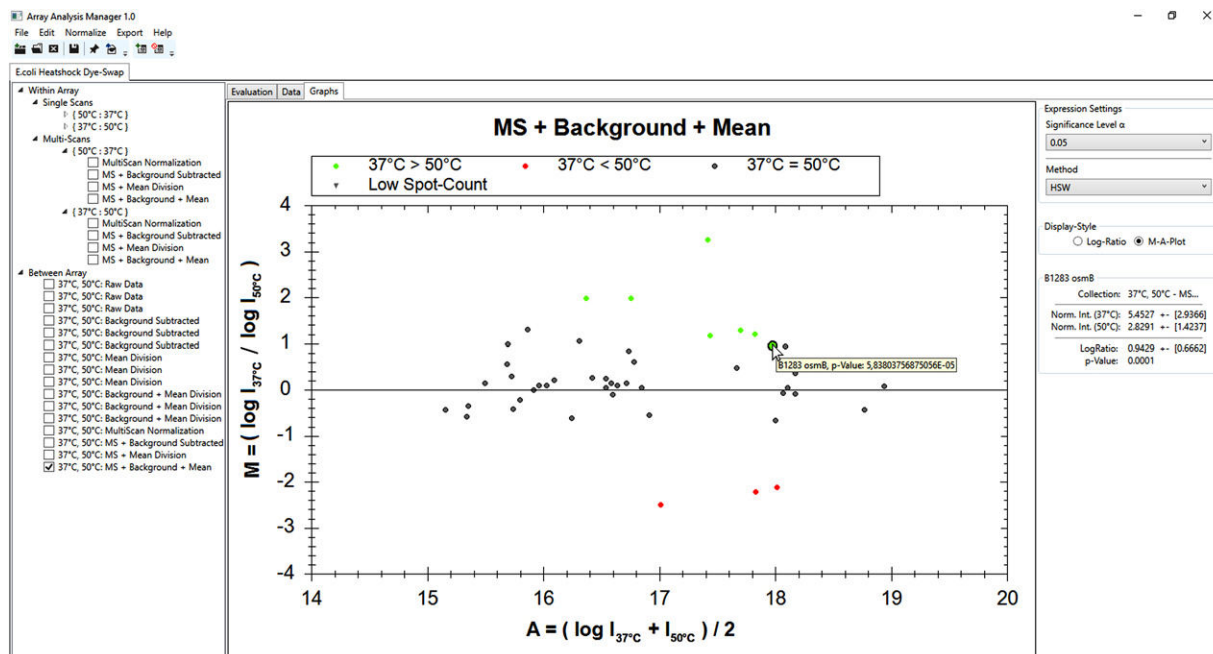
A simplified usability is worthless if the applied methods are not adaptable to the data situation and their effect is not transparently evaluated. This is why this tool was equipped with a normalization evaluation algorithm, which gives the experimenter feedback on the effect each applied normalization has on different bias sources.

Finally, the possibility to export results as widely used data formats allows for a quick, simplified submission, contributing to the expansion and acceleration of scientific exchange.

The whole analysis process is guided by a tooltip system. It consists of multiple graphical user interfaces that explain every step of the analysis process and provides the user with background information and can be turned on and off individually (Figs. 2 and 3). Functions are included that allow the quick exchange of project data. The possibility of saving the results as an .xlsx document (and in GEO format) streamlines the data processing and publishing process.



**Figure 2.** Screenshot of the New Project/New Experiment tooltip. The tooltip system is designed to guide the experimenter through the analysis of his data and provides useful information and links to more detailed information. Screenshot taken on a Windows 10 system.



**Figure 3.** Exemplary screenshot of the graphical user interface of the *Array Analysis Manager*. The “Graphs”-tab allows for easy access to the generated results. The experimenter can display differently normalized feature collections in parallel to compare them. Screenshot taken on a Windows 10 system.



## 4 Concluding remarks

The Array Analysis Manager is an all-in-one program for DNA microarray data analysis. It simplifies the whole process from data processing and experimental design to filtering and data normalization/correction all the way to publishing the results. New methods such as multiscan normalization are included. In the case of photobleaching, this program is the first to incorporate a normalization method. With tools such as the photobleaching bias recognition, the ANOVA-based normalization evaluation, and the tooltip system, the experimenter is guided through the process and empowered to make informed decision regarding the best fitting processing and normalization path for his dataset. All in all, the Array Analysis Manager is a potent software designed to improve microarray analysis transparency, reproducibility, and promote the successful application of the microarray technology.

### Practical application

This software was designed to simplify the microarray analysis process and help the experimenter to make informed decisions about the analysis process to contribute to reproducibility and comparability. Single and multiarray experiments can be processed as well as single and multiscan datasets, allowing for the usage of enhanced dynamic intensity ranges. It provides the experimenter with the first photobleaching recognition and normalization algorithms. Also the analysis of variance (ANOVA) based normalization evaluation provides a vital feedback for the experimenter regarding the individual benefits and/or disadvantages of all possible supported normalization techniques and their combinations. The possibility to export analysis results in Microsoft Excel document (.xlsx) and Gene Expression Omnibus (GEO) formats promotes easy and fast data processing and management. All in all, this software is suited to increase the comparability of microarray experiments by simplifying and harmonizing the data analysis process.

*The authors have declared no conflicts of interest.*

## 5 References

- [1] Ehrenreich, A., DNA microarray technology for the microbiologist: An overview. *Appl. Microbiol. Biotechnol.* 2006, 73, 255–273.
- [2] Spielbauer, B., Stahl, F., Impact of microarray technology in nutrition and food research. *Mol. Nutr. Food Res.* 2005, 49, 908–917.
- [3] Allison, D. B., Cui, X., Page, G. P., Sabripour, M., Microarray data analysis: From disarray to consolidation and consensus. *Nat. Rev. Genet.* 2006, 7, 55–65.
- [4] Microarray Quality Control Consortium, MAQC-II: Analyze that!. *Nat. Biotechnol.* 2010, 28, 761–761.
- [5] Tseng, G. C., Ghosh, D., Feingold, E., Comprehensive literature review and statistical considerations for microarray meta-analysis. *Nucleic Acids Res.* 2012, 40, 3785–3799.
- [6] Harrison, A., Binder, H., Buhot, A., Burden, C. J. et al., Physico-chemical foundations underpinning microarray and next-generation sequencing experiments. *Nucleic Acids Res.* 2013, 41, 2779–2796.
- [7] Shi, L., Perkins, R. G., Fang, H., Tong, W., Reproducible and reliable microarray results through quality control: Good laboratory proficiency and appropriate data analysis practices are essential. *Curr. Opin. Biotechnol.* 2008, 19, 10–18.
- [8] Wu, Z., A review of statistical methods for preprocessing oligonucleotide microarrays. *Stat. Methods Med. Res.* 2009, 18, 533–541.
- [9] Saeys, Y., Inza, I., Larranaga, P., A review of feature selection techniques in bioinformatics. *Bioinformatics* 2007, 23, 2507–2517.
- [10] Drăghici, S., *Statistics and Data Analysis for Microarrays Using R and Bioconductor*, 2nd edition, Taylor & Francis, Boca Raton, FL, USA 2011.
- [11] Gupta, R., Auvinen, P., Thomas, A., Arjas, E., Bayesian hierarchical model for correcting signal saturation in microarrays using pixel intensities. *Stat. Appl. Genet. Mol. Biol.* 2006, 5, Article 20.
- [12] Bengtsson, H., Jonsson, G., Vallon-Christersson, J., Calibration and assessment of channel-specific biases in microarray data with extended dynamical range. *BMC Bioinformatics* 2004, 5, 117–193.
- [13] Khondoker, M. R., Glasbey, C. A., Worton, B. J., Statistical estimation of gene expression using multiple laser scans of microarrays. *Bioinformatics* 2006, 22, 215–219.
- [14] Ambroise, J., Bearzatto, B., Robert, A., Macq, B., Gala, J. L., Combining multiple laser scans of spotted microarrays by means of a two-way ANOVA model. *Stat. Appl. Genet. Mol. Biol.* 2012, 11, Article 8.
- [15] Williams, A., Thomson, E. M., Effects of scanning sensitivity and multiple scan algorithms on microarray data quality. *BMC Bioinformatics* 2010, 11, 127–138.
- [16] Lyng, H., Badiie, A., Svendsrud, D. H., Hovig, E., Myklebost, O., Stokke, T., Profound influence of microarray scanner characteristics on gene expression ratios: analysis and procedure for correction. *BMC Genomics* 2004, 5, 10–19.
- [17] Alston, M. J., Seers, J., Hinton, J. C., Lucchini, S., BABAR: An R package to simplify the normalisation of common reference design microarray-based transcriptomic datasets. *BMC Bioinformatics* 2010, 11, 73–85.
- [18] Fujita, A., Sato, J. R., Ferreira, C. E., Sogayar, M. C., GEDI: A user-friendly toolbox for analysis of large-scale gene expression data. *BMC Bioinformatics* 2007, 8, 457–463.
- [19] Hicks, S. C., Irizarry, R. A., quantro: A data-driven approach to guide the choice of an appropriate normalization method. *Genome Biol.* 2015, 16, Article 117.
- [20] Edgar, R., Domrachev, M., Lash, A. E., Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 2002, 30, 207–210.
- [21] Singh, K. S., Thual, D., Spurio, R., Cannata, N., SaDA: From sampling to data analysis—an extensible open source infrastructure for rapid, robust and automated management and analysis of modern ecological high-throughput microarray data. *Int. J. Environ. Res. Public Health* 2015, 12, 6352–6366.

- [22] Lindner, P., Stahl, F., Statistical evaluation of Microarray experiments. Available at: [https://www.tci.uni-hannover.de/chip\\_bio.html?&L=1](https://www.tci.uni-hannover.de/chip_bio.html?&L=1) (accessed February 28, 2017).
- [23] Microsoft. NET Framework 4.5. Available at: <https://www.microsoft.com/de-de/download/details.aspx?id=30653> (accessed February 17, 2017).
- [24] Microsoft Visual Studio Express. 2012. Available at: <https://blogs.msdn.microsoft.com/visualstudio/2012/06/08/visual-studio-express-2012-for-windows-desktop/> (accessed February 17, 2017).
- [25] Shi, L., Tong, W., Su, Z., Han, T. et al., Microarray scanner calibration curves: Characteristics and implications. *BMC Bioinformatics* 2005, 6 (Suppl 2), S11.
- [26] Repenning, C., *Doctoral Thesis: Statistische Auswertung von Microarray-Daten*, Leibniz Universität Hannover 2010.
- [27] Stennett, E. M. S., Ciuba, M. A., Levitus, M., Photophysical processes in single molecule organic fluorescent probes. *Chem. Soc. Rev.* 2014, 43, 1057–1075.
- [28] Dar, M., Giesler, T., Richardson, R., Cai, C. et al., Development of a novel ozone- and photo-stable HyPer5 red fluorescent dye for array CGH and microarray gene expression analysis with consistent performance irrespective of environmental conditions. *BMC Biotechnol.* 2008, 8, 86–93.
- [29] Satterfield, M. B., Lippa, K., Lu, Z. Q., Salit, M. L., Microarray scanner performance over a five-week period as measured with Cy5 and Cy3 serial dilution slides. *J. Res. Natl. Inst. Stand. Technol.* 2008, 113, 157–174.
- [30] Staal, Y. C. M., van Herwijnen, M. H. M., van Schooten, F. J., van Delft, J. H. M., Application of four dyes in gene expression analyses by microarrays. *BMC Genomics* 2005, 6, 101–113.
- [31] von der Haar, M., Preuß, J. A., von der Haar, K., Lindner, P. et al., The impact of photobleaching on microarray analysis. *Biology* 2015, 4, 556–572.
- [32] Otto, M., *Chemometrie: Statistik und Computereinsatz in der Analytik*, VCH Verlagsgesellschaft GmbH, Weinheim, Germany 1997.

## 4 Conclusion

Many great questions and challenges of this day and age are to be solved only by comprehending and mastering complex dynamic systems. This is true for the incomplete understanding of the systematic connections of genetic and metabolic regulation in molecular biology. Challenges of similar nature can also be found in fields such as neuro science, image processing and robotics (facial recognition, autonomous driving, etc.), climate science (predictive climate and weather modelling) and many more. As a consequence, problems nowadays are most often not confined to a single scientific and engineering discipline, implying that a single discipline's set of tools and methods may not provide an appropriate solution. It is therefore mandatory for researchers and engineers of any kind to obtain and apply interdisciplinary knowledge. On the other hand, several disciplines have developed analogous methods to solve analogous problems. However, these methods might differ not only in nomenclature but also agreed upon criteria and thresholds of significance. All this complicates the interdisciplinary discourse as findings are often not presented in transparent and comprehensible ways, depending on who tries to interpret them.

This thesis contributes to the simplification of scientific discourse regarding DNA microarray based gene expression analysis by improving the comparability, reproducibility and comprehensibility of DNA microarray result data. On the experimental level, obstructing effects that introduce bias into microarray datasets were characterized. Statistical as well as technical normalization solutions were proposed that do not only successfully normalize these effects but do so in a way that allows for their ubiquitous applicability. Additionally, a software was developed that simplifies the microarray data workflow, incorporates new and well established methods, and provides the experimenters with the needed information and guidelines, empowering them to optimize their dataset regardless of the experimenters scientific background and/or prior knowledge.

Future development should advance the photobleaching-recognition ANN to allow the recognition of bleaching for whole genome arrays as well as a broader variety of low density arrays. The ANOVA-based normalization evaluation tool should be refined to evaluate additional bias sources. While commercial analysis suites concentrate on normalizations for their own proprietary whole genome arrays, methods and tools for the analysis and normalization of individually manufactured arrays, often used in research environments, are scarce. Further development should therefore concentrate on advancing the toolset for these setups. Nevertheless new normalization methods that process bias specific to commercial arrays should be implemented to increase the usability for users predominantly working with commercial arrays, e.g. in clinical research and diagnosis. While the recent version of the *Array Analysis*

## Conclusion

---

*Manager* does recognize invalid experimental setups and warns the user, the implementation of an additional experiment planning module would be advisable. An easy-to-use tool that allows an experimenter to plan their experimental design with regard of statistical power and expected significance of the results would be an important step towards standardized and comprehensibly experimental design. The *Array Analysis Manager* was developed using the Microsoft® .NET-framework and WPF. This paves the way for the straightforward implementation of the analysis suite as a XAML-based web-service. Experimenters worldwide would profit from the expanded accessibility of this software, as now prior installation on their own systems is required. Given the agreement of the experimenters, their data could in turn be used to optimize the photobleaching recognition ANN as well as any future ANN to come.

---

## 5 Literature

1. Heller, M.J., DNA microarray technology: Devices, systems, and applications. *Annual Review of Biomedical Engineering* **2002**, *4*, 129-153.
2. Allison, D.B.; Cui, X.; Page, G.P.; Sabripour, M., Microarray data analysis: From disarray to consolidation and consensus. *Nature reviews. Genetics* **2006**, *7*, 55-65.
3. Spielbauer, B.; Stahl, F., Impact of microarray technology in nutrition and food research. *Mol. Nutr. Food Res.* **2005**, *49*, 908-917.
4. Ryden, P.; Andersson, H.; Landfors, M.; Naslund, L.; Hartmanova, B.; Noppa, L.; Sjostedt, A., Evaluation of microarray data normalization procedures using spike-in experiments. *BMC bioinformatics* **2006**, *7*, 300.
5. Thilakarathne, P.J.; Verbeke, G.; Engelen, K.; Marchal, K., A nonlinear mixed-effects model for estimating calibration intervals for unknown concentrations in two-color microarray data with spike-ins. *Statistical applications in genetics and molecular biology* **2009**, *8*, Article 5.
6. Shi, L.; Tong, W.; Su, Z.; Han, T.; Han, J.; Puri, R.K.; Fang, H.; Frueh, F.W.; Goodsaid, F.M.; Guo, L., *et al.*, Microarray scanner calibration curves: Characteristics and implications. *BMC bioinformatics* **2005**, *6 Suppl 2*, S11.
7. Vora, G.J.; Meador, C.E.; Anderson, G.P.; Taitt, C.R., Comparison of detection and signal amplification methods for DNA microarrays. *Mol Cell Probes* **2008**, *22*, 294-300.
8. Tjong, V.; Yu, H.; Hucknall, A.; Rangarajan, S.; Chilkoti, A., Amplified on-chip fluorescence detection of DNA hybridization by surface-initiated enzymatic polymerization. *Anal Chem* **2011**, *83*, 5153-5159.
9. Ricco, R.; Meneghello, A.; Enrichi, F., Signal enhancement in DNA microarray using dye doped silica nanoparticles: Application to human papilloma virus (hpv) detection. *Biosensors & bioelectronics* **2011**, *26*, 2761-2765.
10. Enrichi, F.; Riccò, R.; Meneghello, A.; Pierobon, R.; Cretaio, E.; Marinello, F.; Schiavuta, P.; Parma, A.; Riello, P.; Benedetti, A., Investigation of luminescent dye-doped or rare-earth-doped monodisperse silica nanospheres for DNA microarray labelling. *Optical Materials* **2010**, *32*, 1652-1658.
11. Li, L.; Li, X.; Li, L.; Wang, J.; Jin, W., Ultra-sensitive DNA assay based on single-molecule detection coupled with fluorescent quantum dot-labeling and its application to determination of messenger rna. *Analytica chimica acta* **2011**, *685*, 52-57.
12. Rousserie, G.; Sukhanova, A.; Even-Desrumeaux, K.; Fleury, F.; Chames, P.; Baty, D.; Oleinikov, V.; Pluot, M.; Cohen, J.H.; Nabiev, I., Semiconductor quantum dots for multiplexed bio-detection on solid-state microarrays. *Critical reviews in oncology/hematology* **2010**, *74*, 1-15.
13. Li, X.; Gao, J.; Liu, D.; Wang, Z., Developing oligonucleotide microarray-based resonance light scattering assay for DNA detection on the pamam dendrimer modified surface. *Analytical Methods* **2010**, *2*, 1008.

## Literature

14. Algar, W.R.; Massey, M.; Krull, U.J., The application of quantum dots, gold nanoparticles and molecular switches to optical nucleic-acid diagnostics. *Trac-Trends Anal. Chem.* **2009**, *28*, 292-306.
15. Russell, S.M., Lisa Ann; Russell, Roslin R., *Microarray technology in practice*. 1st ed.; Elsevier/Academic Press: Amsterdam, Boston, 2009.
16. Staal, Y.C.M.; van Herwijnen, M.H.M.; van Schooten, F.J.; van Delft, J.H.M., Application of four dyes in gene expression analyses by microarrays. *BMC Genomics* **2005**, *6*.
17. Rao, A.N.; Rodesch, C.K.; Grainger, D.W., Supporting information of real-time fluorescent image analysis of DNA spot hybridization kinetics to assess microarray spot heterogeneity (vol 84, pg 9379, 2012). *Anal. Chem.* **2013**, *85*, 4199-4199.
18. Rao, A.N.; Rodesch, C.K.; Grainger, D.W., Real-time fluorescent image analysis of DNA spot hybridization kinetics to assess microarray spot heterogeneity. *Anal. Chem.* **2012**, *84*, 9379-9387.
19. Rao, A.N.; Vandencastele, N.; Gamble, L.J.; Grainger, D.W., High-resolution epifluorescence and time-of-flight secondary ion mass spectrometry chemical imaging comparisons of single DNA microarray spots. *Anal. Chem.* **2012**, *84*, 10628-10636.
20. Vogelsang, J.; Kasper, R.; Steinhauer, C.; Person, B.; Heilemann, M.; Sauer, M.; Tinnefeld, P., A reducing and oxidizing system minimizes photobleaching and blinking of fluorescent dyes. *Angewandte Chemie* **2008**, *47*, 5465-5469.
21. von der Haar, M.; Preuß, J.A.; von der Haar, K.; Lindner, P.; Scheper, T.; Stahl, F., The impact of photobleaching on microarray analysis. *Biology* **2015**, *4*, 556-572.
22. Stennett, E.M.S.; Ciuba, M.A.; Levitus, M., Photophysical processes in single molecule organic fluorescent probes. *Chem. Soc. Rev.* **2014**, *43*, 1057-1075.
23. Song, L.L.; Hennink, E.J.; Young, I.T.; Tanke, H.J., Photobleaching kinetics of fluorescein in quantitative fluorescence microscopy. *Biophys. J.* **1995**, *68*, 2588-2600.
24. Dar, M.; Giesler, T.; Richardson, R.; Cai, C.; Cooper, M.; Lavasani, S.; Kille, P.; Voet, T.; Vermeesch, J., Development of a novel ozone- and photo-stable hyper5 red fluorescent dye for array cgh and microarray gene expression analysis with consistent performance irrespective of environmental conditions. *BMC biotechnology* **2008**, *8*, 86.
25. Satterfield, M.B.; Lipka, K.; Lu, Z.Q.; Salit, M.L., Microarray scanner performance over a five-week period as measured with cy5 and cy3 serial dilution slides. *J. Res. Natl. Inst. Stand. Technol.* **2008**, *113*, 157-174.
26. Branham, W.S.; Melvin, C.D.; Han, T.; Desai, V.G.; Moland, C.L.; Scully, A.T.; Fuscoe, J.C., Elimination of laboratory ozone leads to a dramatic improvement in the reproducibility of microarray gene expression measurements. *BMC biotechnology* **2007**, *7*.
27. Tecan Trading AG, Powerscanner™ – features. [http://lifesciences.tecan.com/products/microarray\\_products/powerscanner/features](http://lifesciences.tecan.com/products/microarray_products/powerscanner/features) (03.11.2017),
28. Kuang, C.; Luo, D.; Liu, X.; Wang, G., Study on factors enhancing photobleaching effect of fluorescent dye. *Measurement* **2013**, *46*, 1393-1398.

## Literature

- 
29. Widengren, J.; Chmyrov, A.; Eggeling, C.; Lofdahl, P.A.; Seidel, C.A.M., Strategies to improve photostabilities in ultrasensitive fluorescence spectroscopy. *J. Phys. Chem. A* **2007**, *111*, 429-440.
  30. van der Velde, J.H.; Oelerich, J.; Huang, J.; Smit, J.H.; Hiermaier, M.; Ploetz, E.; Herrmann, A.; Roelfes, G.; Cordes, T., The power of two: Covalent coupling of photostabilizers for fluorescence applications. *The journal of physical chemistry letters* **2014**, *5*, 3792-3798.
  31. Forster, T., \*zwischenmolekulare energiewanderung und fluoreszenz. *Annalen Der Physik* **1948**, *2*, 55-75.
  32. Dosremedios, C.G.; Moens, P.D.J., Fluorescence resonance energy-transfer spectroscopy is a reliable ruler for measuring structural-changes in proteins - dispelling the problem of the unknown orientation factor. *J. Struct. Biol.* **1995**, *115*, 175-185.
  33. Rao, A.N.; Grainger, D.W., Biophysical properties of nucleic acids at surfaces relevant to microarray performance. *Biomaterials science* **2014**, *2*, 436-471.
  34. Sabanayagam, C.R.; Eid, J.S.; Meller, A., Using fluorescence resonance energy transfer to measure distances along individual DNA molecules: Corrections due to nonideal transfer. *J. Chem. Phys.* **2005**, *122*.
  35. Dinant, C.; van Royen, M.E.; Vermeulen, W.; Houtsmuller, A.B., Fluorescence resonance energy transfer of gfp and yfp by spectral imaging and quantitative acceptor photobleaching. *Journal of Microscopy* **2008**, *231*, 97-104.
  36. Molecular Devices, LLC., Genepix® file formats. [http://mdc.custhelp.com/app/answers/detail/a\\_id/18883/~genepix%C2%AE-file-formats#gpr](http://mdc.custhelp.com/app/answers/detail/a_id/18883/~genepix%C2%AE-file-formats#gpr) (03/12/2017),
  37. Repenning, C. Statistische auswertung von microarray-daten. Leibniz Universität Hannover, 2010.
  38. Lyng, H.; Badiie, A.; Svendsrud, D.H.; Hovig, E.; Myklebost, O.; Stokke, T., Profound influence of microarray scanner characteristics on gene expression ratios: Analysis and procedure for correction. *BMC Genomics* **2004**, *5*.
  39. Drăghici, S., *Statistics and data analysis for microarrays using r and bioconductor, second edition*. Taylor & Francis: 2011.
  40. Landfors, M.; Fahlen, J.; Ryden, P., Mc-normalization: A novel method for dye-normalization of two-channel microarray data. *Statistical applications in genetics and molecular biology* **2009**, *8*, Article 42.
  41. Sistare, F.D.; Chen, J.J.; Morris, S.M.; Domon, O.E.; Pine, P.S.; Rosenzweig, B.A., Dye bias correction in dual-labeled cDNA microarray gene expression measurements. *Environmental Health Perspectives* **2004**.
  42. Dombkowski, A.A.; Thibodeau, B.J.; Starcevic, S.L.; Novak, R.F., Gene-specific dye bias in microarray reference designs. *Febs Letters* **2004**, *560*, 120-124.
  43. Bengtsson, H.; Jonsson, G.; Vallon-Christersson, J., Calibration and assessment of channel-specific biases in microarray data with extended dynamical range. *BMC bioinformatics* **2004**, *5*.

## Literature

- 
44. Ambroise, J.; Bearzatto, B.; Robert, A.; Macq, B.; Gala, J.L., Combining multiple laser scans of spotted microarrays by means of a two-way anova model. *Statistical applications in genetics and molecular biology* **2012**, *11*, Article 8.
  45. Khondoker, M.R.; Glasbey, C.A.; Worton, B.J., Statistical estimation of gene expression using multiple laser scans of microarrays. *Bioinformatics* **2006**, *22*, 215-219.
  46. Gupta, R.; Auvinen, P.; Thomas, A.; Arjas, E., Bayesian hierarchical model for correcting signal saturation in microarrays using pixel intensities. *Statistical applications in genetics and molecular biology* **2006**, *5*.
  47. Yang, Y.; Stafford, P.; Kim, Y., Segmentation and intensity estimation for microarray images with saturated pixels. *BMC bioinformatics* **2011**, *12*, 462.
  48. Maqc-ii: Analyze that! *Nat. Biotechnol.* **2010**, *28*, 761-761.
  49. Shi, L.; Perkins, R.G.; Fang, H.; Tong, W., Reproducible and reliable microarray results through quality control: Good laboratory proficiency and appropriate data analysis practices are essential. *Current opinion in biotechnology* **2008**, *19*, 10-18.
  50. Webb, P.M.; Merritt, M.A.; Boyle, G.M.; Green, A.C., Microarrays and epidemiology: Not the beginning of the end but the end of the beginning.... *Cancer Epidemiology Biomarkers & Prevention* **2007**, *16*, 637-638.
  51. Pan, K.-H.; Lih, C.-J.; Cohen, S.N., Effects of threshold choice on biological conclusions reached during analysis of gene expression by DNA microarrays. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 8961-8965.
  52. Westfall, P.H.; Lin, Y.L.; Young, S.S., *Resampling-based multiple testing*. Sas Inst Inc: Cary, 1990; p 1359-1362.
  53. Tusher, V.G.; Tibshirani, R.; Chu, G., Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 5116-5121.
  54. Saeys, Y.; Inza, I.; Larranaga, P., A review of feature selection techniques in bioinformatics. *Bioinformatics* **2007**, *23*, 2507-2517.
  55. Kerr, M.K.; Afshari, C.A.; Bennett, L.; Bushel, P.; Martinez, J.; Walker, N.J.; Churchill, G.A., Statistical analysis of a gene expression microarray experiment with replication. *Stat. Sin.* **2002**, *12*, 203-217.
  56. Kerr, M.K.; Martin, M.; Churchill, G.A., Analysis of variance for gene expression microarray data. *J. Comput. Biol.* **2000**, *7*, 819-837.
  57. Khan, S.; Greiner, R., Finding discriminatory genes: A methodology for validating microarray studies. **2013**, 64-71.
  58. Clark, N.R.; Hu, K.S.; Feldmann, A.S.; Kou, Y.; Chen, E.Y.; Duan, Q.; Ma'ayan, A., The characteristic direction: A geometrical approach to identify differentially expressed genes. *BMC bioinformatics* **2014**, *15*.
  59. Agilent Technologies, Genespring gx. <http://www.genomics.agilent.com/en/Microarray-Data-Analysis-Software/GeneSpring-GX/?cid=AG-PT-130&tabId=AG-PR-1061> (3.21.2017),



## Literature

- 
60. Illumina Inc., Genomestudio software.  
<https://www.illumina.com/techniques/microarrays/array-data-analysis-experimental-design/genomestudio.html> (3.21.2017),
  61. Tseng, G.C.; Ghosh, D.; Feingold, E., Comprehensive literature review and statistical considerations for microarray meta-analysis. *Nucleic acids research* **2012**, *40*, 3785-3799.
  62. Berger, B.; Peng, J.; Singh, M., Computational solutions for omics data. *Nat. Rev. Genet.* **2013**, *14*, 333-346.
  63. Asyali, M.H.; Colak, D.; Demirkaya, O.; Inan, M.S., Gene expression profile classification: A review. *Curr. Bioinform.* **2006**, *1*, 55-73.
  64. Liu, J.; Pham, T.D., Fuzzy clustering for microarray data analysis: A review. *Curr. Bioinform.* **2011**, *6*, 427-443.
  65. Li, Z.J.; Yang, A.; Chen, X.; Zeng, L.J.; Cao, T., A composite method for feature selection of microarray data. *J. Comput. Theor. Nanosci.* **2014**, *11*, 472-476.
  66. Love, T.; Carriquiry, A., Repeated measurements on distinct scales with censoring-a bayesian approach applied to microarray analysis of maize. *J. Am. Stat. Assoc.* **2009**, *104*, 524-540.
  67. Gust, A.; Zander, A.; Gietl, A.; Holzmeister, P.; Schulz, S.; Lalkens, B.; Tinnefeld, P.; Grohmann, D., A starting point for fluorescence-based single-molecule measurements in biomolecular research. *Molecules* **2014**, *19*, 15824-15865.
  68. Fujita, A.; Sato, J.R.; Ferreira, C.E.; Sogayar, M.C., Gedi: A user-friendly toolbox for analysis of large-scale gene expression data. *BMC bioinformatics* **2007**, *8*, 457.
  69. Hicks, S.C.; Irizarry, R.A., Quantro: A data-driven approach to guide the choice of an appropriate normalization method. *Genome Biology* **2015**, *16*.
  70. Singh, K.S.; Thual, D.; Spurio, R.; Cannata, N., Sada: From sampling to data analysis-an extensible open source infrastructure for rapid, robust and automated management and analysis of modern ecological high-throughput microarray data. *Int. J. Environ. Res. Public Health* **2015**, *12*, 6352-6366.
  71. Selvaraj, S.; Natarajan, J., Microarray data analysis and mining tools. *Bioinformation* **2011**, *6*, 95-99.
  72. Otto, M., Chemometrie: Statistik und computereinsatz in der analytik. In *Chemometrie: Statistik und computereinsatz in der analytik*, VCH Verlagsgesellschaft mbH: Weinheim (Germany), 1997.

## 6 Figures

Figure 2.1 Schematic model of a typical DNA microarray experiment workflow.....	4
Figure 2.2: Schematic model of Cy3-fluorescence-labelling based DNA microarray scan imaging.	5
Figure 3.1: Graphical abstract of „The Impact of Photobleaching on Microarray Analysis“ .....	15
Figure 3.2: Graphical abstract of „Optimization of Cyanine Dye Stability and Analysis of FRET Interaction on DNA Microarrays” .....	34
Figure 3.3: Graphical abstract of “Array Analysis Manager – An automated DNA microarray analysis tool simplifying microarray data filtering, bias recognition, normalization and expression analysis” .....	51

---

## 7 Tables

<b>Table 3.1:</b> Calculations of sums of squares (SS) for a linear regression's analysis of variance (ANOVA). $y$ - dependent variable; $j$ - vector containing all $y$ -means of the $i^{\text{th}}$ measurement/observation; $f$ - number of factor combinations; $n$ - number of measurements/observations; $p$ - number of independent parameters; $y = 1ni = 1nyi$ - total mean [72] .....	52
<b>Table 9.1:</b> Contributions .....	69

## 8 Licenses

The publications found in chapters 3.1 and 3.2 were published in open access journals. The publication found in chapter 3.3 was published in “Engineering in Life Sciences”. In agreement with the respective spokespersons these publications are included in this thesis. These spokespersons further confirmed that licenses are not necessary.

## 9 Contributions

The publications presented in chapter 3 were created in cooperation with other scientist who contributed guidance, analysis and/or practical work. All contributors are cited within the publications but additionally listed in **Table 9.1**.

**Table 9.1:** Contributions

<b>Publication</b>	<b>Contributing Authors</b>
"The Impact of Photobleaching on Microarray Analysis"	Marcel von der Haar (70%)
	John-Alexander Preuß (10%)
	Kathrin von der Haar (5%)
	Dr. Patrick Lindner (5%)
	Dr. Frank Stahl (5%)
"Optimization of Cyanine Dye Stability and Analysis of FRET Interaction on DNA Microarrays"	Prof. Thomas Scheper (5%)
	Marcel von der Haar (70%)
	Christopher Heuer (5%)
	Kathrin von der Haar (5%)
	Martin Pähler (5%)
"Array Analysis Manager – An automated DNA microarray analysis tool simplifying microarray data filtering, bias recognition, normalization and expression analysis"	Dr. Patrick Lindner (5%)
	Dr. Frank Stahl (5%)
	Prof. Thomas Scheper (5%)
	Marcel von der Haar (80%)
	Dr. Patrick Lindner (10%)
	Dr. Frank Stahl (5%)
	Prof. Thomas Scheper (5%)

## 10 Author's Resume

---

### Personal Details

Name, Surname: von der Haar, Marcel Philipp  
 Birth Name: Koch  
 Date of Birth: 26.05.1988  
 Nationality: Germany

### Academic Qualifications

since 01/2014 *Institut für Technische Chemie, Leibniz Universität Hannover*  
 Doctorate: Dr. rer. nat. (Chemistry)

10/2011 - 10/2013 *Leibniz Universität Hannover*  
 Life Science (M. Sc.), Grade: 1.3

10/2008 - 09/2013 *Leibniz Universität Hannover*  
 Life Science (B. Sc.), Grade: 1.2

08/2000 - 07/2007 *Kranich Gymnasium Salzgitter*  
 Abitur/A-Levels/High School Certificate, Grade: 1.5

### Professional Experience

08/2012 - 03/2013 *Institut für Technische Chemie der Leibniz Universität Hannover*  
 Student Assistant

02/2012 - 04/2012 *Alexander Silberman Institute of Life Sciences,  
 Hebrew University of Jerusalem*  
 Internship as Student Assistant

09/2010 - 03/2012 *Institut für Siedlungswasserwirtschaft und Abfalltechnik,  
 Leibniz Universität Hannover*  
 Student Assistant

07/2008 *Helmholtz Zentrum für Infektionskrankheiten, Braunschweig*  
 Internship as Laboratory Assistant

10/2007 - 06/2008 *Environmental Protection Authority Salzgitter*  
 Civilian Service

### Academic Distinctions

03/2013 - 08/2013 Scholarship of the *Ernest-Solvay-Stiftung* for the Manufacturing of International Research Projects (Master's Thesis at *Colorado State University, CO, USA*)

2012 Scholarship of the *German Technion Society* for a Internship in Israel

2010 -2012 Annual Scholarship of the *Federal State of Lower Saxony* (twice)