Improving Experimental Outcomes in Kinome Microarrays Through Quality Control

A thesis submitted to the
College of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements
for the degree of Master of Science
in the Department of Computer Science
University of Saskatchewan
Saskatoon

By
Connor Denomy

Permission to Use

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Disclaimer

Reference in this thesis to any specific commercial products, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement, recommendation, or favoring by the University of Saskatchewan. The views and opinions of the author expressed herein do not state or reflect those of the University of Saskatchewan, and shall not be used for advertising or product endorsement purposes.

Requests for permission to copy or to make other uses of materials in this thesis in whole or part should be addressed to:

Head of the Department of Computer Science
176 Thorvaldson Building, 110 Science Place
University of Saskatchewan
Saskatoon, Saskatchewan S7N 5C9 Canada

OR

Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan S7N 5C9 Canada

Abstract

Peptide microarrays consisting of defined phosphorylation target sites are an effective approach for high-throughput analysis of cellular kinase (kinome) activity. Kinome peptide arrays are highly customizable and do not require species-specific reagents to measure kinase activity, making them amenable for kinome analysis in any species. However, the data emerging from experiments with kinome peptide arrays exhibit a large amount of variability. To mitigate this issue, we introduce PIIKA 2.5 to expand upon existing software by providing three important quality control features in an aim to increase the accuracy and consistency of kinome results, which often suffer due to the aforementioned variability. The first feature concerns the size of the virtual circle drawn around each probe in microarray analysis software (spot size). This circle creates the boundary between pixels interpreted as foreground signal and pixels interpreted as background signal. In this thesis, it is shown that too large of a spot size creates abnormal data characteristics, such as high skewness (the asymmetry of the distribution of the data), that can alter downstream results. Here, a feature is presented that alerts users to the existence of improper spot size and informs them of the need to perform a manual alignment to enhance the quality of the raw intensity data, based on the skewness of the data as determined by examination of the mean and median of each dataset. The second feature uses inter-array comparisons to identify outlier arrays that sometimes emerge as a consequence of technical or unknown issues. The work shown in this thesis indicates that the removal of said outlier arrays improves downstream analysis and interpretation. The third feature is a new background correction method, background scaling. Here, it is demonstrated to sharply reduce spatial biases in comparison to the most popular background correction method, background subtraction. Collectively, the modifications presented in PIIKA 2.5 allow users to identify low-quality data, improve clustering of treatment groups, reduce unintended effects, and enhance reproducibility in kinome analysis. The web-based and stand-alone versions of PIIKA 2.5 are freely accessible at `http://saphire.usask.ca/saphire/piika`.

Contents

List of Tables

List of Figures

List of Abbreviations

ADP	Adenosine 5'-diphosphate
ATP	Adenosine 5'-triphosphate
ANOVA	Analysis of variance
DBI	Davies–Bouldin index
DI	Dunn index
DNA	Deoxyribonucleic acid
INF-γ	Interferon gamma
ISMB	Intelligent Systems for Molecular Biology
ISCB	International Society for Computational Biology
LPS	Lipopolysaccharide
PCA	Principal component analysis
PHOSFER	PHOsphorylation Site FindER
PIIKA	Platform for Intelligent, Integrated Kinome Analysis
RNA	Ribonucleic acid
RSE	Repetitive signal enhancement
SAPHIRE	Saskatchewan Phosphorylation Internet Resource
TLR4	Toll-like receptor 4
TNFa	Tumor necrosis factor alpha
t-SNE	t-distributed stochastic neighbour embedding
VSN	Variance-stabilizing normalization

1 Introduction

In eukaryotic cells, there is no signalling process more important than kinase-mediated phosphorylation. Nearly all cellular processes are regulated by these reversible protein modifications. A total of 538 kinases are known in humans, representing approximately 2% of entire human genome [1]. Phosphorylation interactions are also attractive targets for therapies, as kinase inhibitors are a well-characterized category of drugs with a plethora already approved for use by many regulatory agencies. This is in addition to the fact that the improper regulation and mutation of kinases is involved in many human diseases.

Across the biological sciences, an increasing amount of attention is being paid to a broad category of research known as "omics", wherein an aspect of a cell or cell population is analysed completely. With respect to phosphorylation interactions, this type of investigation is referred to as kinomics. The kinome is defined as the complete set of phosphorylation interactions carried out by kinases. One popular method of investigation is using peptide kinase microarrays, which are specially designed microarrays that utilize peptides with known phosphorylation events to detect the sum of all phosphorylation events in a cell population at a point in time. Kinome profiling using peptide kinase microarrays has shown to be a robust method that does not require species-specific reagents and is amenable to high throughput techniques, giving a significant advantage over other popular methods, such as mass spectrometry [2]. In addition, this approach has an advantage over nucleic acid microarrays for studying gene expression (DNA and RNA microarrays) in its ability to provide insight closer to the process of interest. Kinase interactions precede RNA production (transcription) and changes can be manifested as a response immediately after an exposure to a cell stimulus, allowing for a more precise "snapshot" of the regulatory events directly following stimulation. The kinome profile of a cell is quickly reactive in response to new stimuli, as nearly all cell actions are mediated directly through phosphorylation events.

As with most, if not all, high-throughput approaches, the vast amount of data generated by kinome analysis make it nearly impossible to analyse the data by hand. A software suite, PIIKA [3], was developed to perform this analysis, replacing the use of analysis methods designed for DNA and RNA microarrays. This was then expanded into PIIKA 2 [4], which included deeper statistical analysis and more visualizations. Even with these advancements, obstacles still remain in kinome analysis. Often, these issues are related to low-quality inputs given to PIIKA, rather than the program itself. This thesis aims to overcome some of these obstacles and fix a gap in the current kinome analysis workflow by adding measures that allow for the assessment of quality control issues. It also seeks to identify the sources of these issues and incorporates existing knowledge of quality control methods from nucleic acid microarray analysis.

The overarching structure of this thesis is as follows: Chapter 2 provides the necessary background information from both a biological and computational perspective and concludes with the salient details of a kinase microarray experiment. Following this, in Chapter 3, a description of the goals of this thesis is given. Next, Chapter 4 contains a description of the methodology used in the subsequent sections and the mathematical formulae relevant to this thesis. A two-stage approach is used to discuss the results of this thesis. The first stage, Chapter 5, outlines the preliminary analysis which provides the basis on which the quality control measures were founded and their development. These developments were created as an extension to PIIKA, PIIKA 2.5. Chapter 6 first provides detailed descriptions of the datasets used throughout the rest of the chapter. Following this, the chapter demonstrates the background scaling feature's validation in isolation. Then, it describes the results of applying the developed quality control features to three different datasets for validation. Of these datasets, the first two (Sections 6.3 and 6.4) had some correction performed prior to the analysis presented in this thesis. The third dataset (Section 6.5) demonstrates the use of all three quality control features together. Finally, the last chapter (Chapter 7) discusses the limitations, future directions, and conclusions of this thesis.

Parts of this thesis have been presented to the research community in the forms of a poster (DOI:10.7490/F1000RESEARCH.1118212.1)[5] at the 2020 International Conference on Intelligent Systems for Molecular Biology (ISMB) hosted by the International Society for Computational Biology (ISCB) and as a research article published in PLOS One on September 10, 2021 [6]. Some of the text in this thesis has been adapted from these earlier publications. The background scaling method was originally proposed by Conor Lazarou [7] prior to its implementation and validation in this thesis.

2 Background

This section describes the background information necessary to fully understand this project. Firstly, a brief summary of the basic biochemistry related to the processes of cellular signalling and phosphorylation is presented. This also includes the definition of "the kinome" and the specific advantages of using the kinome for analysis over other forms of high-throughput analyses. Section 2.2 then describes the nature of microarray experiments and the underlying technology. Following this, an explanation and summary of the computational and statistical concepts relevant to kinome microarray analysis is provided, including descriptions of microarrays, microarray analysis technology, clustering, cluster validation, the previous work on the kinase peptide microarray-specific platform PIIKA (Platform for Intelligent, Integrated Kinome Analysis), and descriptions of problems known to affect kinome microarray experiments. Finally, the entire workflow of a kinome experiment is outlined from start to finish, along with definitions of all of the important terminology that will be presented in the rest of this work.

2.1 Biological Background

2.1.1 Cellular Signalling and Protein Phosphorylation

One of the most important functions that proteins execute in cellular biology is cell signalling, the process of turning external cellular stimuli into a physical cellular response. In very general terms (there are many exceptions and elements unique to specific signalling pathways), the process is as follows: First, external stimuli are detected by receptors on the surface of the cell membrane. This can either be via direct binding of a molecule (referred to as a ligand in this context) or in response to environmental conditions. This induces a conformational change in the receptor, promoting the recruitment of signalling proteins. Generally, signalling proteins interact with other signalling proteins, initiating a process known as a signalling cascade. The signalling cascade amplifies the initial signal exponentially, as each successive signalling protein in the chain can interact with multiple downstream signalling proteins. This process typically terminates with the activation of an enzyme that regulates gene expression, resulting in increased or decreased levels of a target protein; the direct activation of a target enzyme or transcription factor to mediate some biological response, producing a cellular effect (i.e., metabolic alterations, initiation of apoptosis); and/or cellular export of molecules for intracellular signalling. In all cases, from start to finish, the entire signalling pathway is almost entirely facilitated through protein–protein interactions. A simplified diagram of cell signalling is shown in Figure 2.1.

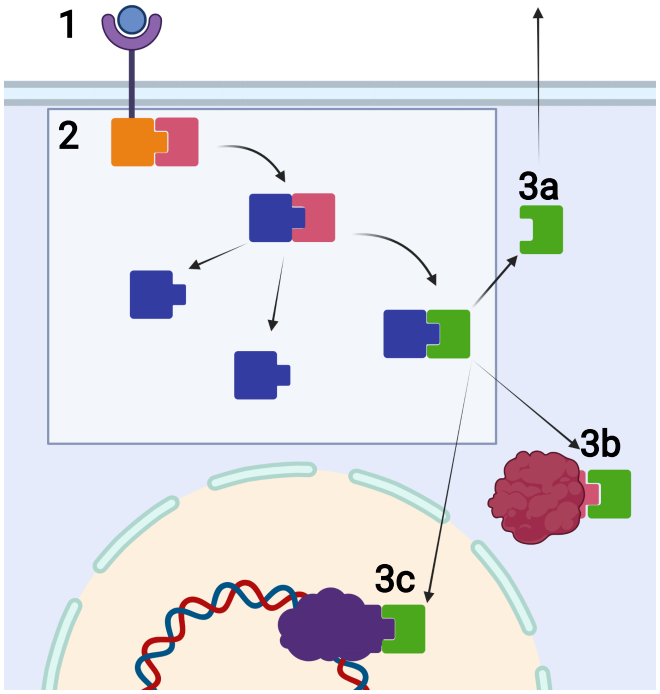

Figure 2.1: Generalized Depiction of Cell Signalling. In cells, external stimuli are detected by receptors on the surface of the cell membrane **(1)**. This initiates a signalling cascade **(2)**. This cascade amplifies in intensity, with each successive step activating more targets than the last, resulting in an exponential increase in signalling activity. The final stage of the signalling cascade involves a terminal protein which can go on to perform a variety of tasks, such as export **(3a)**, direct activation of enzymes **(3b)**, or initiation of gene expression control **(3c)**.

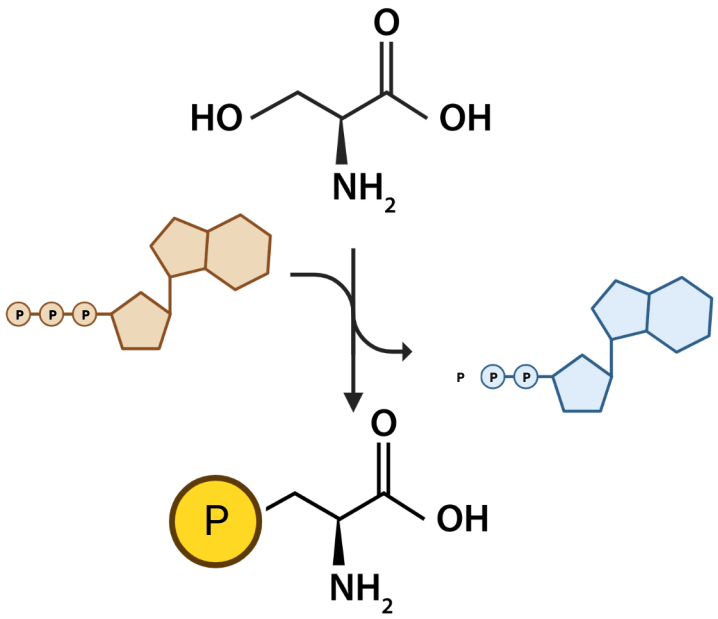

Figure 2.2: Adenosine 5'-triphosphate-mediated Serine Phosphorylation. Simplified diagram of the chemical process of phosphorylation. A serine amino acid is shown before and after modification by the addition of a phosphoryl group donated by adenosine 5'-triphosphate (ATP, shown in orange). In donating the phosphoryl group, the ATP molecule is converted to adenosine 5'-diphosphate (ADP, shown in blue). This serine is shown detached from the protein backbone. The amino acids threonine and tyrosine are also capable of being phosphorylated in a similar manner.

An important category of protein–protein interactions involved in cell signalling is conducted by one specific molecular process: phosphorylation, executed by proteins called kinases. Phosphorylation is a process that uses kinases to move a phosphoryl group from a donor molecule (usually adenosine 5'-triphosphate (ATP)) to the side chain hydroxyl group of a serine, threonine, or tyrosine amino acid (Figure 2.2). Kinase-mediated phosphorylation is a central mechanism of signal transduction and regulation of protein activities in eukaryotes. These reversible protein modifications serve as the defining events that immediately precede many cellular responses, represent highly attractive targets for therapeutic intervention [8], and can affect cellular signalling, resulting in disease [9]. In the context of biology, there are also a number of important factors that limit this interaction, such as steric hindrance, activity of the kinase required for catalysis of the reaction, and blockage created by molecules such as inhibitors, and/or the conformation of the protein with respect to the target phosphorylation site.

2.1.2 Kinome

Given the importance of kinase-mediated phosphorylation, there is increasing priority to define the complete set of kinases and their activity (the kinome) to investigate complex biology, reveal potential therapeutic targets, and identify biomarkers of important phenotypes. This has given rise to a number of technological approaches for defining the kinome [10]. Kinome profiling using peptide microarrays provides a robust and

versatile method with the decided advantages of not requiring species-specific reagents and being amenable to high-throughput analysis.

In peptide microarrays for kinome analysis, surrogate substrates of kinases are represented by short (15-mer) peptides, each containing a phosphoacceptor site in the central position surrounded by the amino acids that are normally present in vivo to best represent the local sequence of the phosphorylation site. Upon exposure to a cellular lysate, the extent to which each peptide is modified through phosphorylation depends upon the activity of the corresponding kinase. By comparing lysates from cells exposed to different biological conditions, it is possible to determine the relative degree of kinase activity, as well as to anticipate the relative extent of modification of the proteins represented by the peptides. Peptide arrays have proven to be a low-cost, versatile tool for kinome profiling. In particular, the opportunity to create customized arrays, coupled with the emergence of software platforms that predict the phosphoproteome of virtually any species [11][12], enables a versatile analysis option, in particular for research where species-specific reagents are limited.

Kinome analysis through kinase target peptide microarrays has two distinct advantages over comparable methods for other "omics" paradigms, such as genome or transcriptome analysis through nucleic acid microarrays. First, it is possible to view the cell function at a level closer to the property at hand. Often, the aim of omics experiments is to determine the biological reality of a cell in question, that is, the functions and pathways that are currently being activated and responsible for presented phenotypes. In essence, the kinome is a "snapshot" of the analysed cell at the point of lysis, revealing exactly which elements of signalling pathways are highly active, unobstructed by interactions which take place at the transcriptional and epigenetic levels. Second, kinases represent excellent drug targets as a result of kinase structure, which has allowed relative ease in the design of kinase inhibitors [13]. Additionally, there is a large catalogue of regulatory agency-approved kinase inhibitors available for purchase from biochemical suppliers to support pharmaceutical investigations of diseases involving kinases.

2.2 Computational and Statistical Background

2.2.1 Microarray Analysis

High-throughput analysis is an approach that exploits the power of modern computers to create an unbiased profile of a large number of results, as opposed to low-throughput techniques, which often analyse only one target of interest at a time. This allows for the detection of multiple activities and novel discoveries in a single assay. One such high-throughput technology is represented by nucleic acid microarrays. Following their first successful applications [14, 15], nucleic acid microarrays have quickly become a leading method of genomic and transcriptional analysis. Typically, microarrays have hundreds to thousands of spots (probes), each representing an individual sample (Figure 2.3). These probes are then organized into blocks (usually identical to act as technical replicates). The use of microarrays was initially limited to specific type of analysis, but has since been greatly expanded for a diversity of tasks beyond nucleic acid analysis [16, 17]. In addition to

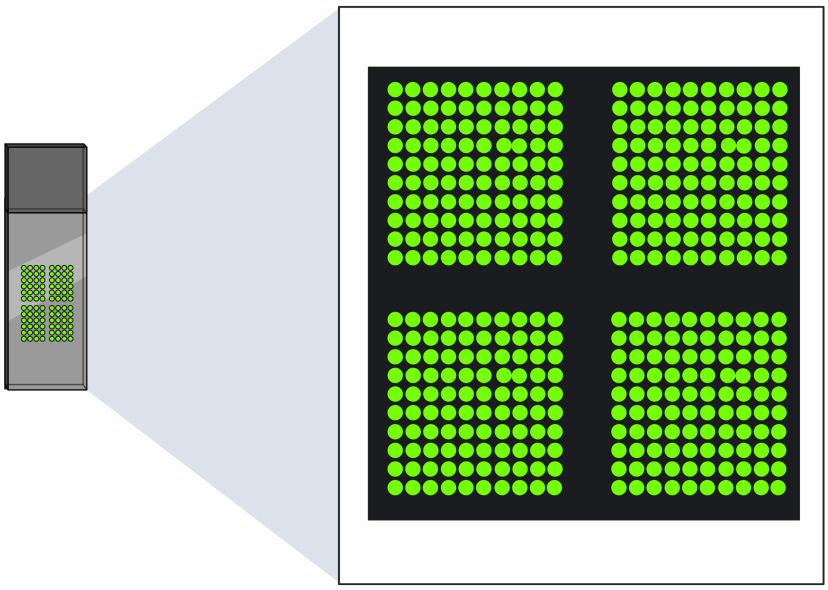

Figure 2.3: Depiction of a Microarray. Microarrays are composed of hundreds to thousands of individual probes that each represent an individual sample printed onto a glass slide. Arrays typically consist of several blocks of probe replicates; the array pictured here has four blocks.

new applications and more advanced nucleic acid microarrays, other types of microarrays have been invented, such those targeted at the high-throughput analysis of carbohydrates [18], lipids [19], antibodies [20], and most important to this thesis, phosphoacceptor peptides. While DNA sequencing and RNA-Seq technology has dramatically improved and superseded nucleic acid microarrays for many applications, these other types of arrays are currently growing in popularity. Despite their growing popularity, these other types of arrays often still rely on the foundational computational methods developed for nucleic acids microarrays.

Initially, it would seem that these computational methods have broad applicability and use across different types of microarrays, such as kinome microarrays. Of course, the standard repertoire of statistical methods including classical hypothesis testing (e.g., Student's t-test), analysis of variance (ANOVA), correction for multiple comparisons (e.g., the Bonferroni correction), correlation analysis (e.g., Pearson or Spearman's rank correlation coefficient), and linear/logistic regression are useful for all microarray technologies. However, past these fundamental statistical operations, bespoke methods may be required for other microarray-based approaches. For example, the numerous methods of data normalization show different levels of applicability for different type of microarrays, as the sources of systematic biases vary between techniques [21]. For the same reasons, there are also several preprocessing techniques available, each with different advantages and varying utility with respect to different types of microarray technology [22, 23]. As a result, it is clear that different types of microarray technologies require dedicated analysis platforms.

2.2.2 Clustering

One approach which has been extensively utilized for analysis in all microarray technologies is clustering. Clustering is an unsupervised machine learning technique that utilizes the values of the data without any a priori knowledge. As defined by Sorin Drăghici in *Data Analysis Tools for DNA Microarrays* [24], "clustering is the process of grouping together similar entities. Clustering can be done on any data: genes, samples, time points in a time series, etc. The particular type of input makes no difference to the clustering algorithm. The algorithm will treat all inputs as a set of n numbers or an n-dimensional vector". The process of a clustering algorithm first involves defining similarity using a distance metric, a formula that calculates how close two points are to each other in n-dimensional space. Following this, there are a multitude of different clustering algorithms with different advantages and use cases [25]. At the end of this process, groups of data points are formed.

Given these groups, it is important to assess the quality of the specific clustering. Drăghici also describes several ways to perform this assessment. The two most important to this work are "the size of the clusters vs. the distance to the nearest cluster" and "the average of the distances between the members of a cluster and the cluster center".

2.2.3 Existing Software for Peptide Array Data

A key event in the advancement of kinome analysis technology was the development of software customized for the biological and technological nuances associated with the peptide array data. The Platform for Integrated, Intelligent Kinome Analysis (PIIKA) provides the ability to identify differential phosphorylation events that are responsive to a particular treatment [26]. PIIKA also enables identification of peptides with inconsistencies among the technical replicates on an individual array as well as among biological replicates (i.e., different responses to the same treatment). The PIIKA platform was later expanded upon with PIIKA 2, which incorporated additional tools in the categories of cluster analysis, statistical analysis, and data visualization [4]. The peptide array technology, with analysis of the data through the PIIKA/PIIKA 2 software platform, has proven to be effective for the investigation of a number of biological processes in a wide range of species [27, 28, 29].

PIIKA also introduced methods for a critical step which has yet to be discussed in this thesis: normalization. Normalization is not relevant to the quality control features presented in this thesis, but nonetheless represents an important tool in quality control. Normalization applies a mathematical function to correct for the differences between arrays so that they can be compared to each other more accurately. Three normalization methods are included in PIIKA: log2 normalization, generalized log2 normalization, and variance-stabilizing normalization (VSN) [3, 30]. Recent developments in normalization methods in the context of kinome profiling include the newly developed repetitive signal enhancement (RSE) normalization, which uses an iterative process to stabilize spots and select only the spots that meet quality control criteria and excludes

spots which are inconsistent across technical replicates [31]. A final note on normalization is that it may be the case that certain types of normalization are better for data with specific characteristics, or that some normalization methods are better than others in all cases. This debate is ongoing, relevant to all forms of high-throughput technology, and is unlikely to stop soon [32, 33, 34].

2.3 Issues Known to Affect Kinome Microarrays

2.3.1 Spatial Bias

In many other microarray-based analysis techniques, variation in intensity associated with the physical location (spatial bias) is a well-known issue. This non-biological bias is very well characterized in nucleic acid arrays and has been known to display different bias patterns (i.e., increasing intensity from one end of the array to the other or from the middle to the edge and vice versa, and unpredictable patterns from dye artifacts) and has several theoretical origins (e.g., dye diffusion and printing error) [35, 36, 37, 38, 39]. Ultimately, this has an effect on experimental reproducibility. Due to the popularity of nucleic acid microarrays, several normalization methods have been developed for their platforms, some of which are over two decades old [40, 41, 42, 43]. Similar trends have recently been reported in antibody microarrays [44] and high-throughput drug screens [45].

In kinome analysis specifically, this bias and its nature have been largely unexplored, but results in an apparent increase in signal towards the bottom of the microarray, resulting in a trend between the position of the probe and its intensity. Though this can be corrected through background correction, the most straightforward method of background correction, background subtraction, is sometimes insufficient at removing the effect of this bias entirely [6].

2.3.2 Temporal Bias

Similar to spatial bias, batch effects have been a well-documented occurrence prevalent in many high-throughput analysis platforms [46]. In a comprehensive review of the subject, Leek et al. defines batch effects as "sub-groups of measurements that have qualitatively different behaviour across conditions and are unrelated to the biological or scientific variables in a study. For example, batch effects may occur if a subset of experiments was run on Monday and another set on Tuesday, if two technicians were responsible for different subsets of the experiments or if two different lots of reagents, chips or instruments were used." [47]. In the same article, they describe the ubiquity in the presence of batch effects throughout published data and online resources/databases. These effects are a powerful source of potential variability in microarray data, and can lead to false positive or false negative results and erroneous conclusions. However, this effect has yet to be extensively investigated in peptide microarrays for kinome analysis. The term "temporal bias", used in this thesis, refers specifically to batch effects associated with the day that the experiment was conducted and has

been observed in kinome analysis experiments.

2.4 Kinome Experiment Workflow and Key Terms

In this section, the entire process of conducting a kinome analysis experiment using PIIKA will be detailed from start to finish to give context to the remainder of this thesis. Additionally, a flow chart is provided as a reference (Figure 2.4). Key terms that are used frequently throughout this work are described here and written in bold font for emphasis and reference.

Initial experimental design begins with formulating a hypothesis and designing an experimental protocol to test the hypothesis. A proper experimental design for a kinome microarray experiment not only includes the animal species, cell types, and treatment conditions, it also includes the process of selecting an appropriate number of biological replicates and technical replicates. **Biological replicates** are identical arrays that use samples taken from a different animal (of the same species) but are otherwise subjected to the same experimental protocol. **Technical replicates** can refer to either replicates of individual peptides repeated on an array or to distinct and separate analyses of the same biological sample. Technical replicates are performed to improve or assess the accuracy of results.

The selection of peptides is often the next step in a kinome microarray experiment. Alternatives to the process of peptide selection exist, such as prefabricated arrays of predetermined peptides. A typical design consists of a mixture of peptides of different predictive origins. These could include control peptides, randomly selected peptides, peptides selected using manual selection based on specific proteins and phosphorylation targets of interest, or putative phosphorylation sites relevant to the research question (as determined based on prior knowledge, literature analysis, online databases [48, 49, 50, 51], GO terms [52, 53]), and computationally predicted peptides, using software such as PHOSFER [54] and DAPPLE [11], both available at saphire.usask.ca, hereafter referred to as **SAPHIRE**. Another related approach, most useful for species from which the protein biochemistry has not been as well characterized, begins with sequences surrounding the phosphorylation sites related to regulatory pathways and processes of interest in more well-characterized species (such as humans or mice). Following this, one can use software such as BLASTp [55] for the prediction of orthologous sequences (different sequences in which the main function has been conserved across species after divergent evolution) in the species of interest [56].

Once the experiment has been designed and the peptides selected, the selected peptides are robotically printed in a grid on a physical microchip, the **microarray**, often by a company such as JPT or Biosynth (Pepscan). From here on, the term "**array**" will be used to refer to either the physical peptide microarray or the experimental intensity data obtained from such an array, while microarray refers exclusively to the physical glass slide on which the peptides are printed. Unless otherwise noted (e.g., "nucleic acid microarray") all usages of these terms refer specifically to kinome peptide microarrays as opposed to any other type.

Upon obtaining the microarray, the experimental procedure is conducted. There are a huge variety of

experimental protocols described in the literature for kinome analysis of different cell/tissue types, species, and conditions, but the general process is as follows: First, the samples of interest are collected. These samples could be from any number of sources and can include bacterial, blood, cell, tissue, and whole-animal samples (e.g., insects). Samples may be exposed to a condition of interest either before collection (in the case of comparing infected vs. uninfected live animals) or after collection (in the case of some tissue and cell experiments). Finally, this process ends with the chemical rupturing (lysis) of cells and collection of the resulting cellular lysate.

After this step, the process of a kinome microarray experiment is much less divergent (more or less similar across experiments). The collected lysate from the previous step is applied to the microarray and kinases present in the lysate selectively phosphorylate the target peptides. Following washing to remove non-specific interactions, **detection** of the levels of phosphorylation is performed. This is commonly performed with a **stain** or **dye** which selectively binds to phosphorylated peptides. Alternative methods of detection also exist, such as using ATP labelled with fluorescent tags which transfer to the target peptide upon phosphorylation. Another alternative is the use of antibodies with fluorescent tags. These antibodies can specifically bind to defined phosphorylated peptide sequences (**phosphopeptides**) or bind to phosphopeptides without sequence specificity. Following application of the detection method, the information on the microarrays is ready to be quantitatively determined.

The first step in collecting the results of a kinome microarray experiment is to place the microarray into a **scanner**, a machine which emits an excitation wavelength of light to activate fluorescence and capture a high-resolution photograph of the microarray. This process can be referred as the "**imaging**" or "**scanning**" step of kinome analysis. This photograph is then imported into software for determination of the intensities of the pixels in the image. As the intensity of fluorescence is directly related to the degree of phosphorylation, the term "**intensity**" is used for both the level of phosphorylation as well as the fluorescent intensity of the pixels as determined by the scanning software. The software that performs this analysis is referred to throughout this work as the "**scanner/scanning software**", or "**vendor scanner software**" when in reference to the proprietary software that is provided by the scanner manufacturer with purchase of the scanner. This process requires an initial **alignment**, which describes the locations of each individual peptide (**probe**) on the microarray to the software. To facilitate analysis and verify technical reproducibility, probes are typically organized into several distinct groups, or **blocks**. The alignment of probes and blocks can be determined given the parameters of the microarray (distance between probes, number of probes, distance between blocks, number of blocks, etc.), automatically detected using a **spot-detecting algorithm**, or manually performed. Around each probe, a virtual circle is drawn (hereafter referred to as a **spot**). By default, the size of these spots is the same for every probe on the microarray, but each spot can be manually adjusted. The final result of this alignment process is a **grid** of spots to analyse.

Using the grid of locations, the fluorescent intensities of the pixels in each spot are measured. In addition to this, the software also **labels** each pixel as being either part of the foreground or background based on the

pixel's location with respect to the spot. Pixels inside the spot are labelled as foreground pixels, while pixels directly outside the spot (within a predefined radius which defaults to the radius of the spot) are labelled as background pixels. The intensity value of an individual pixel can also be referred to as the "**signal**". The output of the software is a file with many summary statistics regarding the characteristics of the array, most notably the mean foreground and mean background intensities for each spot.

The final step of the kinome experimental workflow involves data processing, statistical analysis, and interpretation of the results. This step is detailed specifically as it pertains to using PIIKA 2.5. The initial data processing steps involves conversion of the output of the scanning software to a format more amiable for downstream statistical analysis. This usually involves removal of unnecessary columns and headers, as well as removal of data from probes that are used for scanner software calibration (labelled in the array manifests as "blank-control", "process-control", "blank", "calibration-control", etc.). The scanner software provides users with a file for each individual array. These results must be combined into a single file for use in PIIKA. This has recently been made into a less user-intensive process, hosted at `https://saphire.usask.ca/final_flask/`. Following this, the data files and parameters can be entered into PIIKA. One of the first transformations of the data that PIIKA performs is **background correction**. PIIKA 2.5 currently provides three different options for this: no background correction; **background subtraction**, which is achieved through simple subtraction of the mean background signal from the mean foreground signal of each probe; and **background scaling**, a new method introduced in this work (described in detail in Section 4.1). The last processing step is **normalization**, which applies a mathematical function to correct for the differences between arrays so that they can be compared more accurately.

Based on the input parameters of the user, a number of statistical analyses are performed by PIIKA. It is at this stage that the majority of the content of this work is applicable. Despite the many advantages that analysis of the kinome has, it has a major disadvantage in that there are far fewer tools available for kinome experiments. This work seeks to alleviate that by adding three major quality control improvements to PIIKA. Based on the results of this quality analysis, it may be recommended by PIIKA 2.5 that certain steps of the above process are performed again, and those results rerun in PIIKA 2.5 (see * and †in Figure 2.4).

The output from PIIKA should allow the researcher to determine which peptides on the microarray were the most differentially phosphorylated. From this stage, the researcher can perform additional experimentation to validate the results and apply them as they see fit.

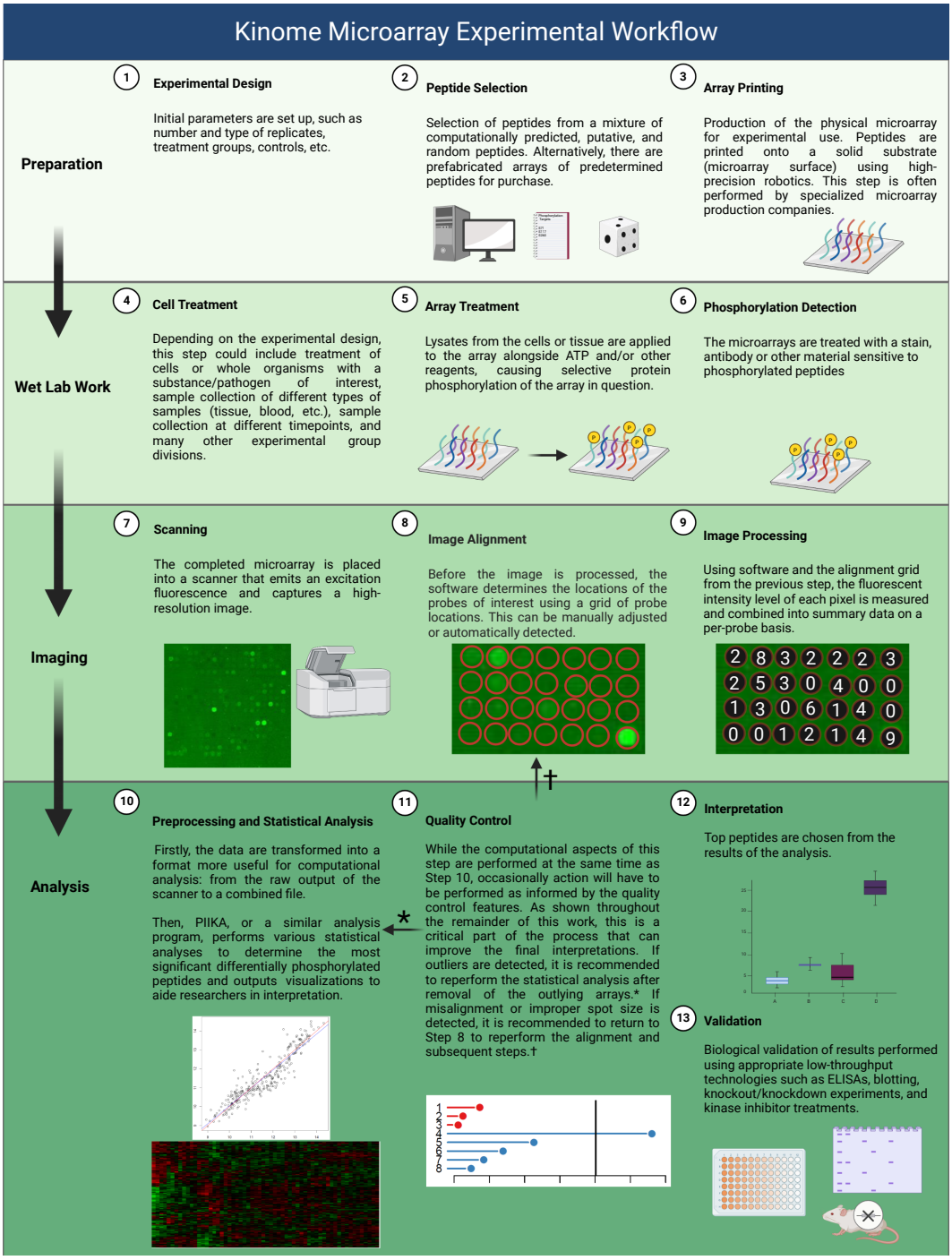

Figure 2.4: Flowchart of the Workflow of a Kinome Experiment The process of performing a kinome experiment as detailed in Section 2.4. *, recommended action if outliers are detected; †, recommended if spot size-induced bias is detected. Adapted from "Multi-step Protocol (Layout 3x4)", by BioRender.com (2022). Retrieved from https://app.biorender.com/biorender-templates.

3 Thesis Objectives

The first major objective of this thesis is to develop an extension to PIIKA 2, called PIIKA 2.5, with three major quality control features that provide mechanisms to enhance the extraction of meaningful biological information from kinome data. The first feature provides a metric by which the mislabelling of pixels during scanning, a serious problem with kinome arrays, can be detected. The second feature provides a metric that allows for the determination of arrays that are significantly deviant from other arrays in the experiment due to unintentional factors (e.g., day of experiment). These arrays have the potential to corrupt downstream analyses, especially analyses in which values are aggregated together. The third feature is a new background correction technique that has the potential to offer better correction of spatial bias than the most common technique, background subtraction. The new utilities are a crucial part of an improved kinome analysis pipeline, and their results are provided to the user alongside the regular output from PIIKA 2.

The second objective of this thesis is to validate these new quality control measures. To this end, it is necessary to characterize other datasets which have specific and apparent quality control issues in addition to datasets with no quality control problems. Each measure requires validation individually, but a combination of all three measures applied to a single dataset is also needed to validate the entire suite of features introduced by PIIKA 2.5.

4 Methodology

4.1 Background Scaling

The background scaling process is performed using the data outputted by the array scanning software, which includes each probe's foreground and background medians. For each probe, its foreground intensity is the total of the intensities of all pixels deemed to be foreground pixels. Similarly, a probe's background intensity is the total of the intensities of all pixels deemed to be background pixels (surrounding the foreground pixels in a predetermined radius). Typically, to compensate for various systematic sources of error related to the location of a spot on the physical array (location bias), foreground intensity is corrected by subtracting the background intensity. However, this does not consider how the local pixel intensities relate to those on the microarray as a whole. The background scaling technique proposed here aims to correct this oversight. With background scaling, both intensity values are first divided by the ratio between the local background intensity and the median background intensity of the entire array (Equation 4.1. This ratio is an additional factor considering the relationship between local background pixels (for the spot under consideration) and the background pixels across the entire array, capturing a normalized estimate of bias at a given location. Following the application of the ratio to both the foreground and background values, the background is then subtracted from the foreground to reduce location-associated bias. As shown in Section 6.2, background scaling sharply reduces the location-associated variation between technical replicates. This method was originally proposed by Conor Lazarou [7, 57].

$$f_s = \frac{f_l}{b_l/b_m} - \frac{b_l}{b_l/b_m} = \frac{f_l(b_m)}{b_l} - b_m \tag{4.1}$$

where f_s is the background-scaled intensity, f_l is the mean foreground intensity, b_l is the mean background intensity, and b_m is the median background intensity.

4.2 Clustering and Cluster Validation Measures

To evaluate the performance of PIIKA 2.5, a method of validation was required. Theoretically (and ideally), the data emerging from identical treatment conditions should have similar characteristics. Based on this assumption, if it is possible to group data together using a clustering algorithm or other means, it is possible to use the quality of grouping to assess performance. Several methods to perform this task were considered: k-means clustering, hierarchical clustering, t-distributed stochastic neighbour embedding (t-SNE), and principal

component analysis (PCA). An emphasis was placed on consistency and reproducibility, as well as ease of computation, understanding, and visualization. While all methods were revisited occasionally to provide small insights when needed, it became clear that dimensionality reduction for visualisation followed by validation was the best approach for these circumstances. While the popularity of t-SNE has been growing quickly since its description in 2008, its stochastic nature was not ideal for this project, especially in the context of outlier analysis. With these considerations, PCA was utilized for dimensionality reduction. Following dimensionality reduction, it is possible to visualize each array in a two-dimensional Cartesian plane, and assess how similar data points are visually and quantitatively using a clustering validation measure. Four different cluster evaluation methods were assessed for their suitability for this task: the Dunn index, the Davies–Bouldin index, the silhouette coefficient, and the Rand index [58]. These are generally divided into two categories: internal cluster evaluation and external cluster validation. Internal cluster validation assesses the quality of clustering based on the intrinsic characteristics of the clustering, rather than any known class labels. In contrast, external cluster validation assesses cluster quality with respect to known class labels.

The Rand index is an external cluster validation measure that calculates the similarity of cluster membership to a set of known class labels. Essentially, it acts as a measurement of the percentage of correct classifications, with a value of 1 representing a perfect classification and a value of 0 representing a completely mismatched classification [59]. The Rand index was the first cluster evaluation measure explored, but was only briefly considered. It was initially believed by the author that an external validation measure would be necessary for this task; however, during the process of validation it was determined that an external validation measure would not be as useful for this application as internal cluster validation measures. External clustering indexes are generally more useful for comparison of clustering algorithms themselves, rather than this type of validation application [24].

The silhouette index is an internal cluster validation measure that evaluates, for each clustered point, the distance between itself and other clusters and the distance between itself and its own cluster. With this measure, a graphical representation can also be created that allows for easy interpretation of results [60]. The silhouette index calculates a silhouette coefficient between -1 and 1, with values closer to 1 indicating that the clusters are more clearly defined and further from other clusters. For unknown and ultimately unexplored reasons, the silhouette index was often very low (near or below 0) and incongruous with both other external cluster evaluation measures as well as visual inspection of the PCA plot (i.e., high silhouette index with visually apparent poor clustering and vice versa). This effect is likely due to the fact that the silhouette index is highly sensitive to both noise and outlying datapoints [22]. Given the high-noise context of this validation experiment and that there are outliers in the data by design, this index was determined to not be applicable for this task.

The Dunn index is an internal cluster validation measure that is based on the distances between pairs of points that are in the same cluster divided by the distances between pairs of points that are in different clusters. The Dunn index is given by Equation 4.2:

$$DI = \frac{min(\delta(C_i, C_j))}{max(\Delta(S_k, S_l))} \tag{4.2}$$

where C_i and C_j are any points within different clusters and $\delta(C_i, C_j)$ is the distance between the points C_i and C_j. Therefore, $min(\delta(C_i, C_j))$ is the smallest distance between any two points of different clusters, or the shortest inter-cluster distance. S_k and S_l are points within the same cluster, and $\Delta(S_k, S_l)$ is the distance between the two points. Therefore, $max(\Delta(S_k, S_l))$ is the largest distance between any two points in the same cluster, or the width of the largest cluster at its widest point. The ratio of these two values calculates the Dunn index, which can be used to compare different clustering algorithms or different ways of using a clustering algorithm with any number of dimensions or clusters, though increases of these values has a strong detrimental effect on computation time. A high Dunn index (close to 1) indicates that the clusters are well separated, while a low Dunn index (close to 0) means that the clusters are not well separated.

The Davies–Bouldin index is another internal cluster validation measure widely used to evaluate the stability and the quality of a clustering algorithm, defined as the ratio of the average of the intra-cluster scatter(the sum of all within-cluster distances) to the distance from the cluster center to the nearest cluster center of the other clusters. In this manner, it is similar to the Dunn index, but uses summary statistics of the inter- and intra- cluster distances in place of maximization/minimization. The Davies–Bouldin index is given by Equation 4.3:

$$DBI = \frac{1}{N} \sum_{i=1}^{N} \max_{j \neq i} \frac{S_i + S_j}{d(C_i, C_j)} \tag{4.3}$$

where N is the number of clusters, S_i is the sum of the intra-cluster scatter for cluster i (the sum of all within-cluster distances), $d(C_i, C_j)$ is the distance between the centroids of clusters i and j, and C is the clustering result. The lower the value of the Davies–Bouldin index, the better the clustering result. Both the Dunn index and the Davies–Bouldin index produced reasonable and consistent results in the context of the cluster validation and showed agreement with qualitative visual assessments of cluster quality, while both the silhouette index and the Rand index gave results which were less useful for the evaluation of results than the Dunn index and the Davis–Bouldin index. Review of the literature also revealed that both indexes are also utilized for similar applications [61, 62, 24]. Therefore, these two methods were selected as a dual measurement for the validation scheme in this thesis.

4.3 Determination of Improper Spot Size

The initial investigation into problems in kinome microarrays revealed a pattern where the automated alignment provided by the scanner vendor's software occasionally produced inconsistent and inaccurate alignments. This pattern does not appear in the literature in the context of nucleic acid microarrays, so it is possible to infer that this problem is unique to kinome microarrays and the software designed for interpretation of nucleic acid microarrays.

The automatic alignment of microarray spots implemented by the GenePix software is often inaccurate on kinome arrays. The vendor scanner software first creates a feature of an estimated size around each spot. It defaults to a feature size that will capture even the largest spots on the array (Figure 4.1A). Pixels of the image inside the feature are labelled as foreground and pixels outside the area of the feature are labelled as background. Using the automatic alignment can cause overestimation of the size of the spot, causing pixels in the spot that should be labelled as background to be labelled as foreground and included with the foreground in subsequent calculations of summary statistics, such as the mean and median. This mislabelling yields a distortion of data that causes the background-corrected mean and median values to diverge from the true values and has major effects on the distribution of intensity values obtained from an array (Figure 4.2). As the spot size becomes larger, more background pixels become mislabelled, rapidly leading to the high-intensity true foreground pixels becoming outnumbered by the low-intensity mislabelled background pixels. This causes the median to shift much lower so that the median point is likely to be the intensity of a mislabelled background pixel, while the mean remains high due to the very high values of the true foreground. This can skew the distribution of background-corrected mean and median intensity values.A more reliable radius of the spot should be smaller, containing less background area. Indeed, it is preferable to make a spot too small as opposed to too large. Additionally, the range of intensity values (difference between the highest and lowest intensity values) is also reduced in data with overestimated feature sizes.

As a simplified example, consider a probe (spot) the size of a single pixel, with a high true foreground intensity of 1000 and a background area surrounding it with an intensity of 1. If the spot size is overestimated to include the background pixels, this changes the value of the median foreground intensity all the way to 1, the value of the true background intensity. The mean is also shifted considerably lower, but not as to the same extent.

In PIIKA 2.5, inaccurate assessment of microarray spot size is detected by utilizing the difference between the median and mean for each spot. Both of these summary statistics are calculated from the intensity values of all foreground pixels deemed to constitute the spot. The mean is not as shifted, however, as the intensities of the most intense, properly labelled pixels are much higher than the intensities of mislabelled pixels. This contrasts with the expected median and mean values (Figure 4.1B). One result of this effect is a skewed distribution of background-corrected mean intensity values and background-corrected median intensity values (Figure 4.2). Another result is that the slope of the median vs. the mean across the entire array is appreciably lower than 1.0 (Figure 4.3). Using this slope as a guideline, we have implemented a three-tiered "stoplight system" to provide feedback to the user. Arrays are given a green, amber, or red light depending on the mean–median slope. Based on preliminary empirical study on hundreds of in-house datasets (results not shown), a mean–median slope of 0.70 or lower is rarely seen in manually corrected arrays. Therefore, a slope of 0.70 is a reasonable threshold for a dataset to be flagged as having unacceptable levels of bias due to over-estimated spot size and in urgent need of alignment correction. A slope less than 0.70 is given a red light, indicating the presence of oversized spot sizes that should be corrected with manual

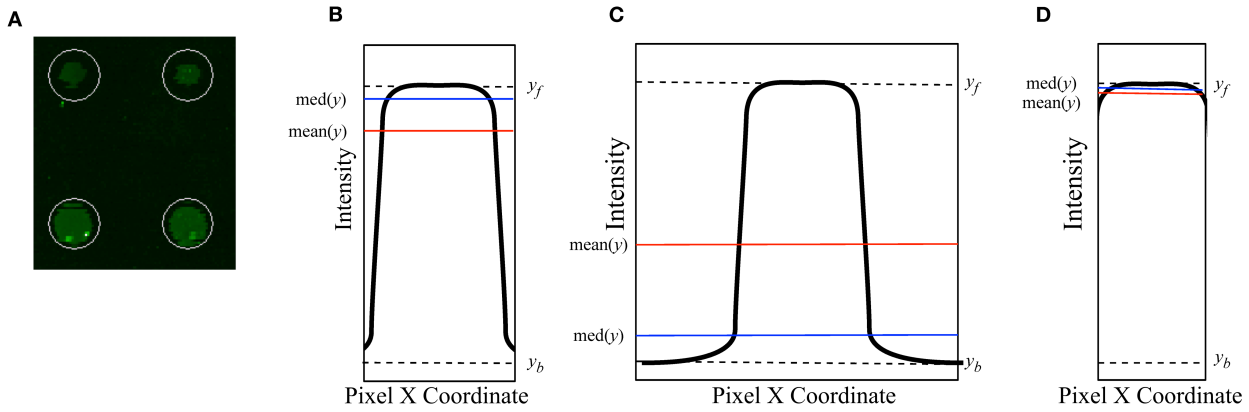

Figure 4.1: Overestimation of Spot Size Introduces Error in Estimation of Mean and Median Spot Intensity. A. The automated spot finding provided by the scanner software defaults to using a feature size (shown by the white circle) that captures the entire area of all spots, even the largest spots on the physical array. **B.** With an accurate feature size, the values of mean(y) and the median (med(y)) give a true representation of the spot. The vertical edges of the rectangle show the boundary between actual foreground and background pixels constituting the spot. **C.** Overestimated feature size, showing the resultant reduced mean (mean(y)) and further reduced median (med(y)). The mean estimated foreground intensity is lowered from the true foreground intensity (y_f), towards the true background intensity (y_b). **D.** Smaller feature size results in increased mean and median, closer to their true values than in panels B or C. Y-axes in panels B, C, and D represent the intensity of the pixel at the position given by the X-axes. The spots in this depiction represents a simplification of a spot that could be created by taking only one pixel per column of the image and does not show that the number of pixels increases quadratically with increasing radius of a circle.

realignment. A slope less than 0.85 (but at least 0.70) is given an amber flag, an indication that manual intervention may not be required, but users should proceed with caution nonetheless. These values are based on the trends seen across hundreds of in-house arrays but are ultimately arbitrary. More analysis is required to determine alternate slope thresholds. In Figure 4.3, the data set shown in panel A2 was given a red light, while that in B2 was given a green light.

This effect leads to two distinct features of the data, most easily observed by examining a plot of the background-corrected median and the background-corrected mean (Figure 4.1C). First, the mislabelled points create a fin-shaped region above the linear regression line, with a background-corrected median intensity less than 5000, and a corrected mean intensity between 1000 and 12000. This feature seems to be a result of oversized spot sizes on high-intensity probes, as the resizing of the probes near the maximal intensity results in their transition into this fin-shaped region. The second effect, caused by the greater lowering of the median than the mean in the poorly aligned probes, is an overall linear regression line with a slope appreciably lower than one, a highly unusual property in comparable microarray technologies and other distributions of biological data. This effect results from the systematic error of probes lower in intensity, and as the low-intensity probes outnumber the high-intensity probes by at least two orders of magnitude in any given array, this effect dominates the summary statistics (mean and median). These elements are demonstrated by the highlighted points in Figure 4.4, in which nine spots were resized for the purpose of demonstration.

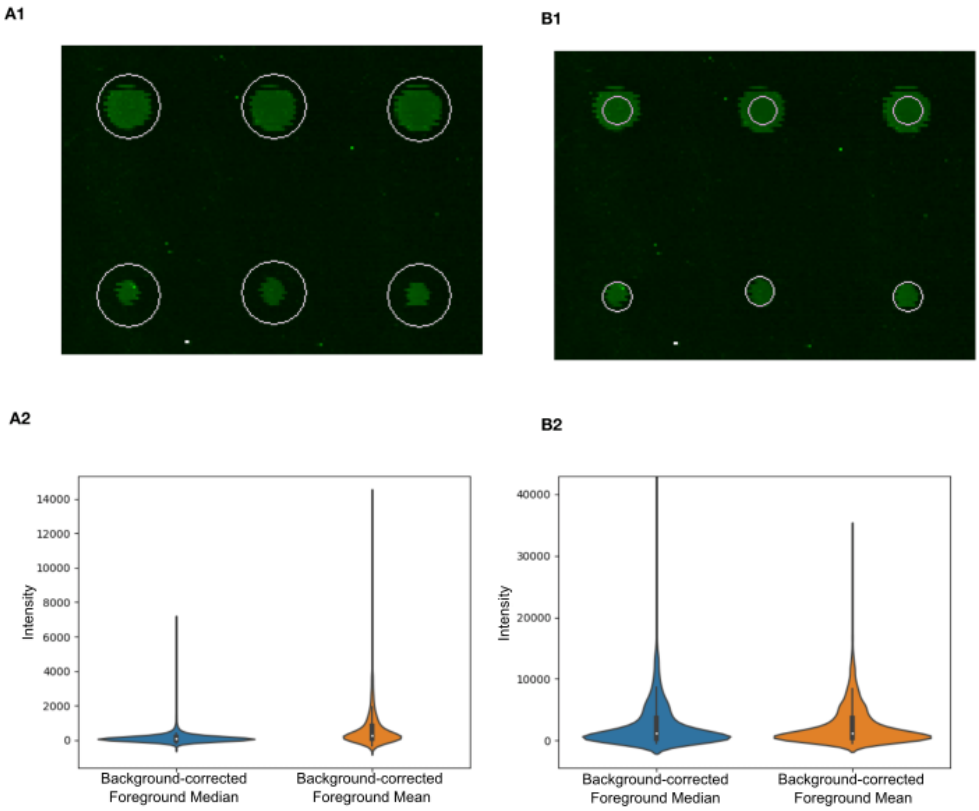

Figure 4.2: Overestimation of Spot Size Can Affect Distribution of Spot Intensity Values. Panel **A1** shows an overestimated feature size as a result of the automated procedure provided by the scanner and scanner software, while panel **B1** shows the same spots subject to a manually adjusted feature size. Panels **A2** and **B2** show violin plots of calculated spot intensities for all spots on the array using the corresponding feature size (A1 and B1, respectively). The overestimated feature size (A1) results in a highly skewed distribution of background-correct median spot intensities and background-corrected mean spot intensities (A2). The effects on the distribution of values includes the range of intensity values, as shown in the y-axis limits shown in Panels A2 and B2. The feature sizes in panels A1 and B1 result in the situations shown in 4.1C and 4.1D, respectively.

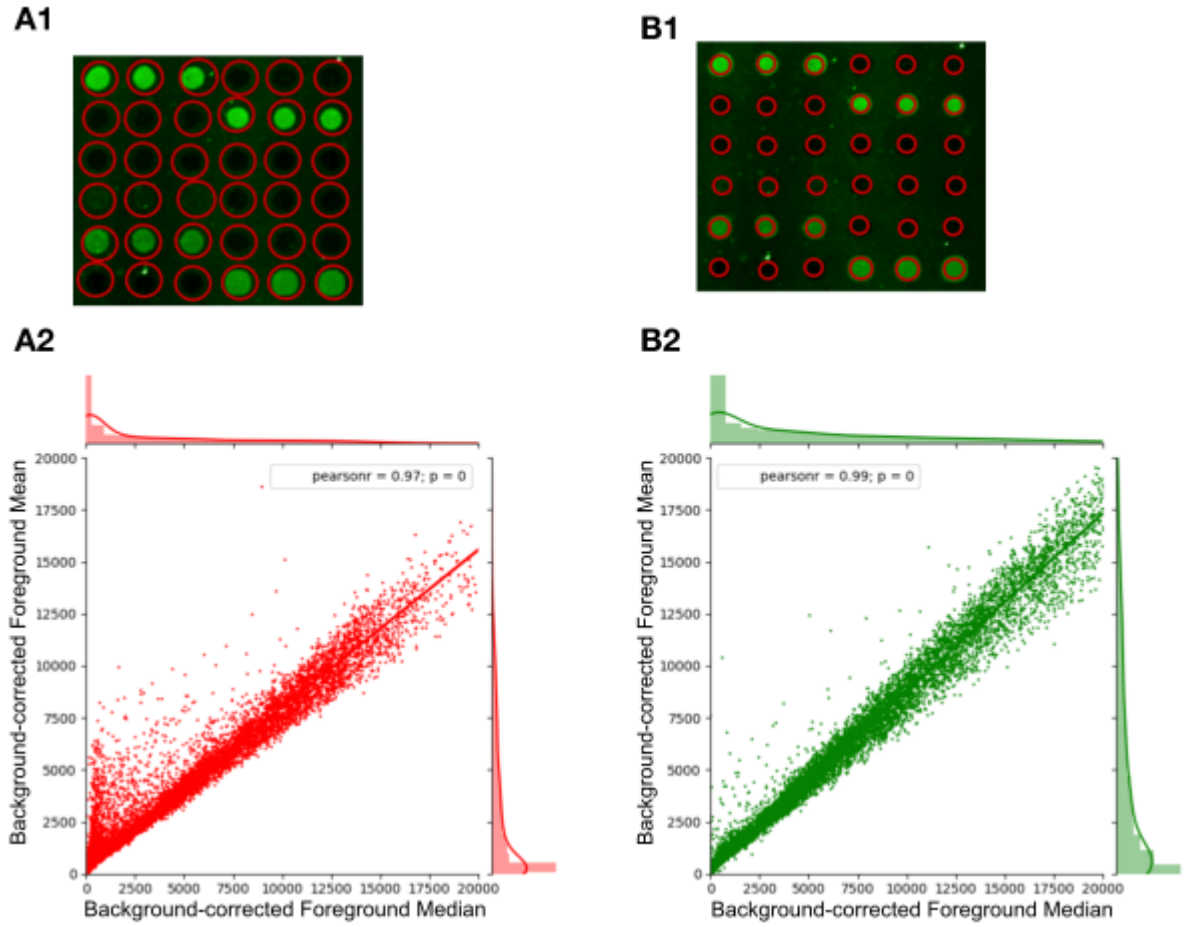

Figure 4.3: Detection of Mislabelled Pixels Due to Errors in Spot Alignment. A1. Scanning grid where spot size is too large. **A2.** Linear regression of foreground means (less background) on foreground medians (less background) of a typical suboptimal spot alignment where the estimated spot (feature) size is much larger than the feature's actual size (shown A1), capturing much of the background area of the array. This leads to the indicative crest of points x <5000 and a slope much lower than 1 (0.755). **B1.** More accurate spot size used for scanning. **B2.** Linear regression of background-corrected foreground means on background-corrected foreground medians of a spot alignment better optimized for kinome microarrays, manually selecting a much smaller diameter for each feature (shown B1). Improved alignment removes the crest feature in the regression plot and raises the slope to 0.855. Frequency histograms along the x- and y-axes show the overall distribution of the background-corrected median and the background-corrected median, respectively. In A1 and B1 actual spots are shown in green, and the boundary of the area labelled as foreground pixels is shown in red. Data are taken from the LPS stimulation experiment. Bright green areas outside the normal spot area are artifacts.

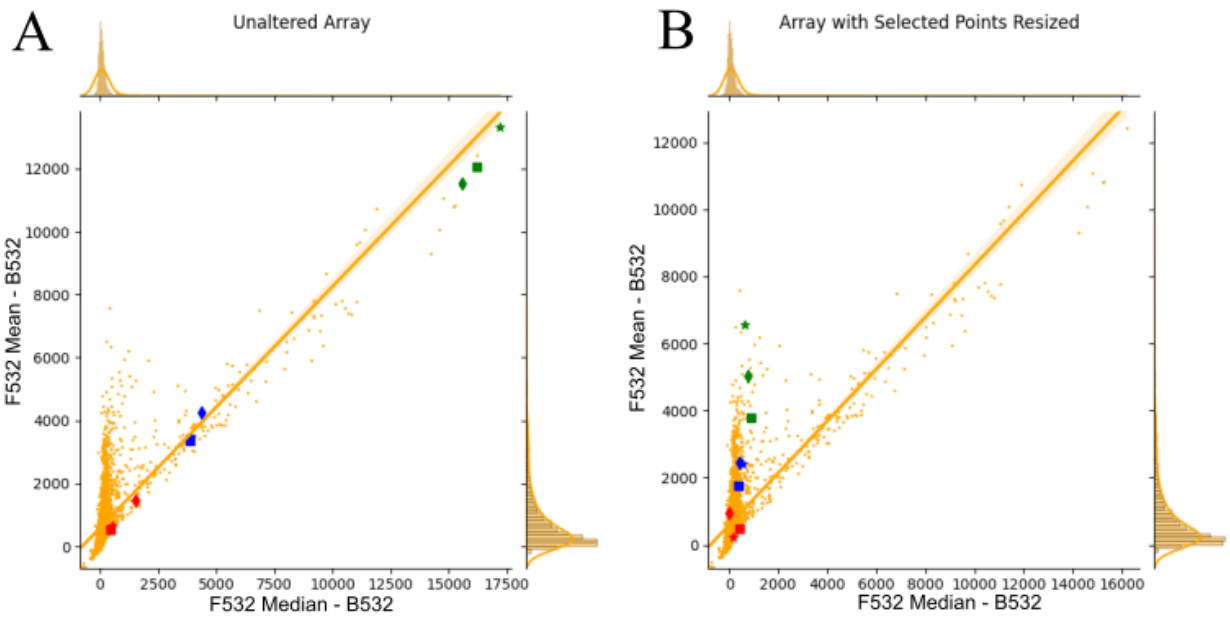

Figure 4.4: Increasing Spot Size Decreases Median More than Mean and has a Stronger Effect on High-Intensity Probes Kinome microarray mean–median plots A. without any alteration and B. with spot size diameter increased from 100px to 300px for the highlighted spots. Three groups of spots were chosen: the first group, shown in green, are three spots selected from those with the highest intensity; the second group, shown in blue, are three spots selected from those with a moderate intensity of approximately 4000; and the third group, shown in red, are spots with lower intensities, one with an intensity of approximately 1500 and two with an intensity near zero. The symbols used for representation have no significance other than to differentiate the individual spots within each group of spots. Spots with higher intensity values are affected much more significantly by an enlarged spot size and the mean is affected far more than the median.

This effect is significantly influenced by the geometric principles inherent to circles, which exaggerates this effect further than what might be expected. Specifically, as the radius of a circle increases linearly, the area of a circle increases quadratically. The relationship between the two values is the well-known equation $A = \pi r^2$. An understanding of this very simple geometric effect in the context of spot size estimation error is critical to understand the significance of this error.

As a final note, there are several scanners and scanning analysis software available. As it was the most practical due to availability of equipment/data, this thesis focuses exclusively on the use of the GenePix scanner produced by the company Molecular Devices and the software that accompanies it. Future investigations may need to examine whether the information presented here is true across other scanners and scanning analysis software or if the proposed approaches are hardware specific. It is possible that these issues are only inherent to this specific combination of hardware and software, and that other programs and/scanners or combinations of the two may not experience this problem. If this is the case, it is the recommendation of this author to switch to a different experimental setup, if possible.

5 Development

5.1 Outlier Detection Methods

This section details the results of preliminary work that influenced the development of the techniques used in this thesis. From the onset, it was clear that specific arrays, for unknown reasons, have unusual characteristics (incongruence of technical replicates, abnormally high/low/noisy background fluorescence, etc.) that could not be precisely quantified or reliably detected. Reviewing the relevant literature from nucleic acid microarrays indicated that an important quality control resource is detection of outlier arrays. As no such methods are incorporated into PIIKA, nor any other kinome-specific analysis platform, the natural next step was to investigate their applicability to kinome microarray data. Three techniques from the package arrayQualityMetrics, a package that is widely used in the analysis of data from nucleotide microarrays, were assessed in terms of their sensitivity and specificity for kinome arrays given several datasets from kinome experiments. The methods were explored using four different datasets: the breast cancer dataset (see section 6.1.2), the bovine macrophage dataset (see Section 6.1.4), the LPS stimulation dataset (with all three detection methods, see Section 6.1.1), and the ileum dataset (see section 6.1.5). A more detailed description of each of those datasets can be found in the methodology section (Section 4). In total, this created a sample size of 164 individual arrays.

5.1.1 Kolmogorov–Smirnov Statistic

The Kolgomorov–Smirnov (KS) test is a widely used method across many statistical and data driven disciplines. This technique can be used to determine if two data samples are drawn from the same underlying distribution. The null hypothesis for the test is that the two data samples are drawn from the same distribution, while the alternative hypothesis is that the two data samples are not drawn from the same underlying distribution. The KS test is a non-parametric test, meaning that it does not make any assumptions about the underlying distribution of the data, but it seems to be the case that kinome microarrays share too much similarity in the distributions for this to be a useful method to detect outliers. In total, just 2 of the 164 arrays (only 1.21% of samples) were determined to be outliers. In this method, outliers are determined as having a distribution that is significantly different than the combined distribution of all arrays.

5.1.2 MA Plots and Hoeffing's D_a

MA plots are widely used in microarray analysis and involve plotting the intensity values of the data on axes of log ratio ($M = log_2(I) - log_2(P)$) and mean average ($A = 1/2log_2(I) + log_2(P)$)), where I is the intensity of the array being analysed and P is intensity of an array composed of the median value (across all arrays in an experiment) of each probe. MA plots provide an overview of the distribution of the intensity values which can then be quantitatively assess with a statistical test such as Hoeffing's D_a. Following this transformation, Hoeffing's D_a is used to compute the independence of M and A, based on the hypothesis that there should be no trend in M as a function of A. This exact protocol has been used throughout the literature to detect the presence of outliers in high-throughput analysis [63, 64].

In assessment for kinome microarrays, this protocol found an outsized portion of microarrays to have significantly high values of Hoeffing's D_a, indicating a low degree of independence between M and A. In this regard, it was found to be not a useful statistic outlier detection for kinome analysis. In total, 73 arrays of the 164 (44.5%) were determined to have a low degree of independence between M and A, including 20 of the 27 arrays from the LPS stimulation (antibody) dataset.

5.1.3 Pairwise Distance

In this method, outlier arrays are determined by inter-array comparisons using measurements of the mean absolute difference between the data for all corresponding probes between arrays. That is, we calculate the distance between each array pair by determining the mean of the absolute values of the differences between all corresponding probes on the two arrays. There is one such mean or distance for each pair of arrays. Then, for each array, the distances between it and all other arrays in the experiment are collected and totalled. Finally, the total distances for all arrays are subjected to a one-way analysis of variance (ANOVA) and p values are calculated. Arrays that are significantly different past a margin of p <0.05 according to Tukey's HSD test, corrected for multiple observations using the Bonferroni correction, are marked as outliers. As in the determination of pixel mislabelling, a "stoplight system" is implemented here as well. Arrays marked as outliers are given a red light and best practice is to remove them entirely. Arrays that trend towards significance (p <0.1) are given an amber light and could be included in subsequent analysis with caution. Other arrays are given a green light by this analysis. One additional calculation is performed. For each array, the sum of distances to all other arrays (i.e., all its inter-array distances) is also determined. The calculation is part of the ANOVA, but the total can be used to represent the variability of each array with respect to all the others. A threshold value is approximated for the value of the sum of distances at which the ANOVA result has a p value of 0.05. This threshold is shown in visualizations.

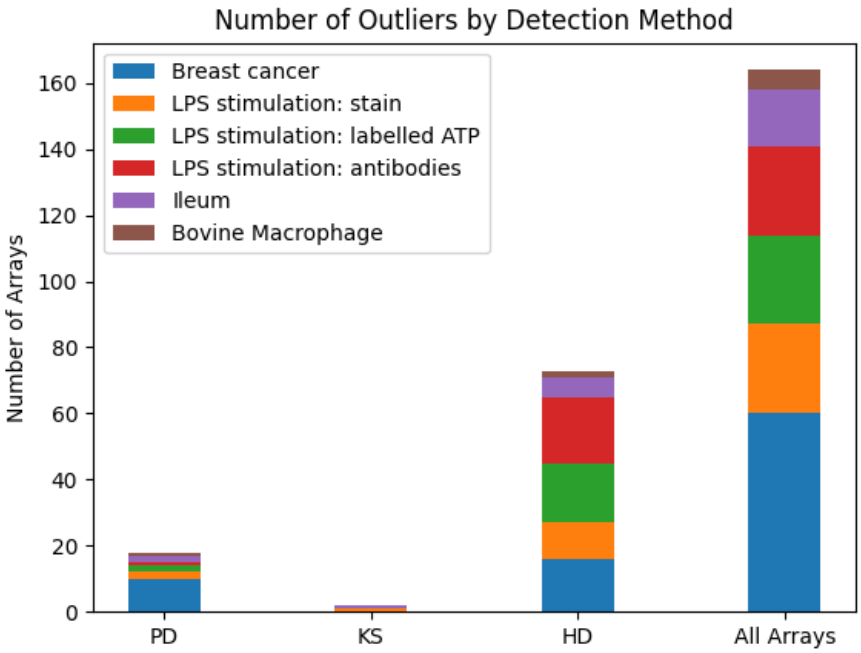

Figure 5.1: Outlier Detection Using Three Different Methods: PD, pairwise distance method; KS, Kolmogorov–Smirnov test; and HD, Hoeffing's D-statistic. Each of the first three bars represent the total number of arrays found to be outliers. The final bar, total arrays, represents the total number of arrays from each analysed dataset. Out of 164 total arrays, the pairwise distance method found 18 (10.97%) to be outliers, the Kolmogorov–Smirnov test found only 2 arrays to be outliers (1.21%), and the Hoeffing's D-statistic method found (44.5%) of arrays to be outliers.

5.1.4 Conclusions

Based on the above analysis, the pairwise distance method appears to be the most applicable technique to analyse kinome data. Out of 164 total arrays, the pairwise distance method labelled 18 (10.97%) arrays as outliers, the Kolmogorov–Smirnov test labelled only 2 arrays as outliers (1.21%), and the MA plot/Hoeffing's D_a method labelled 73 (44.52%) of arrays to as outliers. The three methods had complete overlap; all arrays labelled as outliers by the KS test were also labelled by the pairwise distance method and all arrays labelled as outliers in the pairwise distance method were labelled as outliers but the MA plot/Hoeffing's D_a method. A summary of these results can be seen in Figure 5.1 and in Table 5.1. The pairwise distance method is not overly sensitive, as with Hoeffding's dependence coefficient nor overly specific, as with the Kolmogorov–Smirnov statistic. The results for the LPS stimulation dataset will be discussed further in Sections 6.3 and 6.5. In regard to the bovine macrophage dataset, one array (array 1) was detected to be an outlier with a p value of <0.001. For the ileum dataset, two outliers were detected (arrays 3 and 5), both with p values of less than 0.001.

	BC		LPS-ATP		LPS-Ab		LPS-S		ID		BM		All Arrays	
PD	10	(16.67%)	2	(7.407%)	1	(3.70%)	2	(7.407%)	2	(11.7%)	1	(16.67%)	18	(10.97%)
KS	0	(0%)	0	(0%)	0	(0%)	1	(3.70%)	1	(5.88%)	0	(0%)	2	(1.21%)
HD	16	(26.67%)	18	(66.66%)	20	(74.07%)	11	(40.74%)	6	(35.29%)	2	(33.33%)	73	44.52%
Total	60		27		27		27		17		6		164	

Table 5.1: Summary of Outlier Detection Using Three Different Methods: PD, pairwise distance method; KS, Kolmogorov–Smirnov test; and HD, Hoeffing's D-statistic. Datasets: BC, breast cancer; LPS-ATP, LPS stimulation (radioactivity labelled ATP); LPS-Ab, LPS stimulation (ATP-binding antibodies); LPS-S, LPS stimulation (phosphorylation-specific stain); ID, ileum dataset; BM, bovine macrophage dataset. Thorough descriptions of all datasets described can be found in Section 6.1.

Following this analysis, a chi-squared test for independence was conducted to discover if there was any correlation between treatment group and outlier status. This test failed to prove that there was a statistically significant association between outliers and treatment condition.

As a final note, the repeated reassessment of outliers after outlier removal using these methods (except the KS test) resulted in more arrays being labelled as outliers (i.e., arrays that were not labelled as outliers in the first assessment were labelled as outliers in subsequent assessments after outlier removal). This type of repeated analysis is not recommended as a result. Though users could be given all three test values as an output from PIIKA, it seems highly unlikely, given the results presented here and their subsequent validation, that useful outlier interpretation is to result from the two unused statistics.

5.2 Development Specifications

The original PIIKA program, as well as the PIIKA 2 update, are programmed in R [65] and Perl [66]. The vast majority of the program is implemented in R, including the code that concerns statistical methods and visualization. A large amount of data processing is also performed in R. The Perl files handle the initial data entry and processing, especially in PIIKA 2. PIIKA 2's web-based interface is reliant upon Perl code and the Apache [67] web server for forwarding data to the main program.

PIIKA 2.5 is implemented by adding and modifying the code for PIIKA 2, and uses R for outlier analysis as well as basic data processing. The outlier analysis functionality is adapted from the package arrayQualityMetrics [68], with several modifications and additions. Many aspects of this package, including the generation of the outlier analysis figures, have been repurposed for use in the kinome quality control pipeline of PIIKA 2.5.

In PIIKA 2.5, all processing that is performed outside of R uses Python [69]. Python performs the analysis of data skewness and outputs the visualizations that concern it using aforementioned statistical methods. The non-standard python packages in use in this application are as follows: matplotlib [70], used for visualization of the data; numpy [71], used for its ability to create arrays for data processing and storage; seaborn [72], used to enhance the visualization output from matplotlib; and sklearn [73], used for the creation of the slope using linear regression.

Related to this project, there are several other modifications and upgrades relevant to the user experience of PIIKA 2.5. First, a video tutorial has been created on the usage of PIIKA 2.5, available from the SAPHIRE website at `https://saphire.usask.ca/saphire/piika_beta/piika2.5_user_guide.html`. This was created in response to the comments made by the reviewers of the PIIKA 2.5 manuscript and has become a valuable source of information for new users. Secondly, a feedback page has been created on the SAPHIRE website, allowing researchers experiencing issues to send feedback and get a prompt response.

6 Applications and Validation

While the previous chapter focused on the development of the improvements to PIIKA 2.5, this chapter will discuss the application of the new features of PIIKA 2.5 in pursuit of validation of their effectiveness. First, descriptions and experimental details of the datasets used throughout the remainder of the chapter are provided. Second, the background scaling method will be assessed in isolation for its corrective ability across different experimental variables. Following this, the new quality control features will be assessed through three case studies. Each case study procedure will consist of the application of the quality control improvements and a discussion of the overall results. The first two of these case studies used datasets which had pre-existing quality such that at least one of the new features of PIIKA 2.5 was not necessary. As such, neither of the first two case studies thoroughly exercised the combination of all three of PIIKA 2.5's new quality control measures. Therefore, to completely validate the entire quality control process facilitated by PIIKA 2.5, a third case study will be described to demonstrate the effect of all three major quality control measures in tandem on a single dataset. In all case studies, the effectiveness of these quality control measures will be indicated by an improvement in the grouping of data according to anticipated characteristics as determined by the cluster quality control metrics that are discussed in Section 4.2.

6.1 Datasets Used in this Project

The following sections describe the characteristics of the datasets used throughout this thesis and details regarding their collection.

6.1.1 The LPS Stimulation Dataset

Throughout this thesis, the dataset described in this subsection will be referred to as the "LPS Stimulation dataset".

The utility of the PIIKA 2.5 program was investigated within the context of data from an experiment involving peripheral blood mononuclear cells stimulated with lipopolysaccharide, the ligand for toll-like receptor 4 (TLR4) [74]. This ligand is a part of a well-characterized biological system in this cell population, and provides an established framework for investigation of data analysis tools [56, 75, 76]. A distinct advantage of this system is the rapid induction of the expression and release of tumour necrosis factor alpha (TNFa) in response to LPS stimulation, which provides a readily quantifiable indicator of cell response [77]. The data herein exploit another known property of this system, cortisol's ability to dampen LPS-induced TNF release

[78]. There was a total of 27 arrays in this experiment, representing three biological replicates undergoing three treatments (control, LPS, LPS + cortisol) performed identically on three different days.

Isolation and stimulation of immune cells was performed as follows: Blood was transferred to 50 mL polypropylene tubes and centrifuged at 1400 × g for 30min at 20°C. White blood cells were isolated from the buffy coat and mixed with PBSA + 0.1% EDTA to a final volume of 35mL. The cell suspension was layered onto 15 mL of Ficoll-paque plus (Amersham Biosciences, GH healthcare) and centrifuged at 400 × g for 40min at 20°C. Peripheral blood mononuclear cells (PBMC) from the Ficoll-PBSA interface were collected and resuspended in 50ml cold PBSA +0.1% EDTA. The suspension was centrifuged at 1200rpm for 10 minutes at 4°C to wash. The resultant pellet was resuspended in 50ml cold PBSA (no EDTA) and centrifuged at 1200rpm for 10 minutes at 4°C. This wash was repeated again for a total of three washes and pellets from the same animals were combined. Isolated PBMCs were cultured in RPMI medium (GIBCO) supplemented with 10% heat-inactivated fetal bovine serum. Isolated PBMCs were rested overnight prior to stimulation. Purified PBMCs (10×10^6) were stimulated with 100 ng/mL LPS (Escherichia coli 0111:B4) (Sigma-Aldrich), 100ng/mL LPS with 1µM hydrocortisone (added to media 30 minutes before LPS stimulation), and/or media for 4 hr at 37°C. This quantity and type of LPS was previously shown to induce cellular responses in porcine monocytes. Cells were pelleted and stored at −80°C before use in the peptide arrays. Cell supernatants were diluted (1:2) and used for enzyme-linked immunosorbent assays (ELISAs) for porcine TNFa as per the R&D Systems DuoSet Development Systems ELISA kit.

Application of kinome arrays was performed using previously described protocols [11]. Arrays were manufactured by JPT Innovative Peptide solutions. Peptides were chosen for the arrays based on known phosphorylation sites relevant to the experiment as well as other computationally predicted sites using DAPPLE [11]. A total of 27 arrays were analysed, representing three animals and three different treatments (control, LPS, LPS + cortisol). This was performed in triplicate, with each set of replicates analysed on different days. Each array had a structure consisting of 282 unique peptides with 9 replicates each, arranged into 3 major blocks divided into 16 sections each. Each of the 16 sections per block (48 total) had 7 columns and nine rows. Each block was identical, each containing three replicates of each peptide. The array structure is detailed in Figure 6.1 In addition to the phosphorylation-specific stain, this experiment also used two other methods of detection of phosphorylation: fluorescent-labelled ATP and phosphopeptide-specific antibodies.

Both the data resulting from the phosphorylation-specific stain and the radioactive-labelled ATP were subjects of analysis. The phosphorylation-specific stain data was utilised in the previous publication [6] as well as in Section 6.3. The data from the fluroescent-labelled ATP detection assay was used in Section 6.5 and the experiment structure was identical to that of the phosphorylation-specific stain with the exception of the method of phosphorylation detection. The phosphopeptide-specific antibody data was not used for any analyses as there appeared to be an unusual animal-dependant effect.

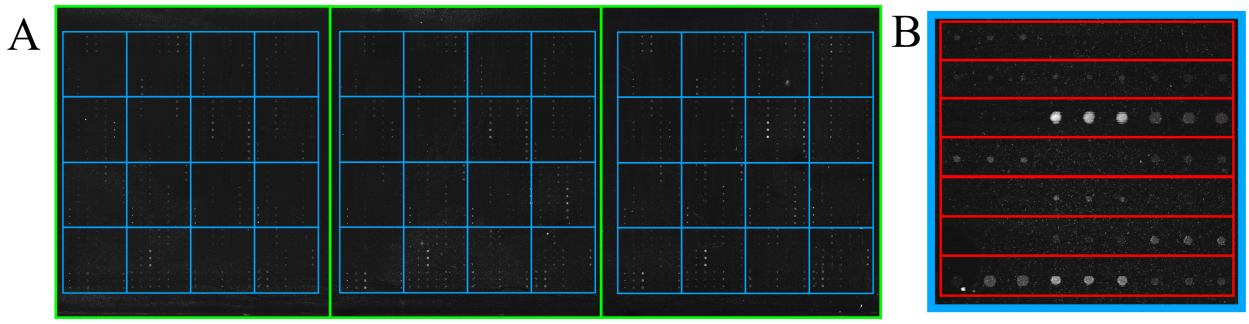

Figure 6.1: Array Layout of the LPS Stimulation Dataset. A. Overarching layout of array structure, detailing the 3 major blocks (green) with 16 divisions (blue). Note that the array is rotated 90 degrees. **B.** Closeup of a single block, showing the 7 rows (red) per block, each with 9 probes. Note that the block has been rotated back to its original orientation.

6.1.2 The Breast Cancer Dataset

The breast cancer dataset is comprised of 60 individual arrays from experiments performed on two successive days (November 16, 2016, and November 17, 2016). Kinome analysis was performed on blood (PBMCs) taken from women with lumps in their breasts, plus healthy controls. Of those with lumps, some were subsequently determined to be cancer and others were non-cancerous. Of the total number of arrays, 20 were arrays of data gathered from experiments using samples from cancerous breast tumours, 20 were data from benign breast tumours, 19 arrays were performed using data from healthy breast tissue samples, and one array was using data from breast papilloma samples. Papillomas are small benign tumours that are highly unlikely to evolve into malignancies [79]. The papilloma sample was considered to be equivalent to the healthy samples in statistical analyses. Isolation of the PBMCs was performed in the same manner as the LPS stimulation dataset.

Each array had a structure consisting of 1440 unique peptides with 9 replicates each, arranged into 3 major blocks divided into 16 sections each. Each of the 16 sections per block (48 total) had 15 columns columns and 18 rows. Each block was identical, each containing three replicates of each peptide. Throughout the remainder of this thesis, this dataset will be referred to as the "breast cancer" dataset.

6.1.3 Other Datasets

This section will describe in brief the characteristics of datasets that were not examined in depth. The datasets described here only played minor roles in this thesis.

6.1.4 Bovine Macrophage Dataset

This dataset was specifically suggested to the author as an that array resulted in poor, inconclusive results. It contains a total of six arrays of data from a single-day experiment on macrophage populations isolated from cow blood, with three samples as controls and three samples treated with interferon gamma (a cytokine

that induces antibody presentation and lysosomal processes in macrophages [80]). This dataset features 341 unique peptides with 9 replicates each, for a total of 3069 peptides. This dataset provides a smaller sample size, which is convenient for its speed of analysis and also was useful for evaluating how outlier detection methods would behave on datasets of different sizes (in contrast to the large breast cancer dataset, for instance). Another relevant aspect of the bovine macrophage dataset is the fact that it contains several probes which returned the mean intensity value "error" from the scanner software, allowing it to act as a useful debugging tool for edge-case testing. This dataset was used exclusively within the context of developing the methodology for outlier detection.

6.1.5 Ileum Dataset

The quality of this specific dataset is exceptional and promising results were obtained from the analysis of this experiment; as such, this array is acting as a control to show how experiments behave ideally. There are 18 arrays in total, three treatment groups with six biological replicates. There is a total of 1440 unique peptides of mixed computationally predicted and manually selected origin. The results of this analysis and experimental details can be found in previously published literature [81]. In brief, the samples represent kinome analysis performed on samples from *M. paratuberculosis*-infected isolated ileal surgical segments and uninfected controls 1 month post-infection. The results showed distinct signalling differences in the interleukin and Wnt-β-catenin pathways. This dataset was used exclusively within the context of developing the methodology of outlier detection.

6.2 Validation of Background Scaling

The first quality control measure developed to improve the experimental results of kinome microarrays was background scaling, a new method of background correction. This section validates the efficacy of that method. This can be performed by demonstrating the ability of background scaling to reduce location bias more effectively than background subtraction. This is demonstrated by applying both background scaling and background subtraction to the LPS stimulation dataset and comparing the frequency of statistically significant correlations between the position of a probe on the physical microarray and its intensity. This is based on the fact that microarrays are generally not designed in a manner that would order probes by expected intensity. As such, the absence of a correlation between position and intensity can be seen as beneficial to the interpretation of results (Figure 6.2). Moreover, a greater reduction in the position–intensity correlation would be evidence of the effectiveness of one method over another.

In this pursuit, background scaling was applied to the raw data of both the fluorescent-labelled ATP and phosphorylation-specific stain detection methods used in the LPS stimulation dataset (Section 6.1.1) to validate the consistency of background scaling across the phosphorylation detection methods. The results of each individual analysis are shown in (Tables 6.1and 6.2), and a summary of the results is provided in Figure 6.3.

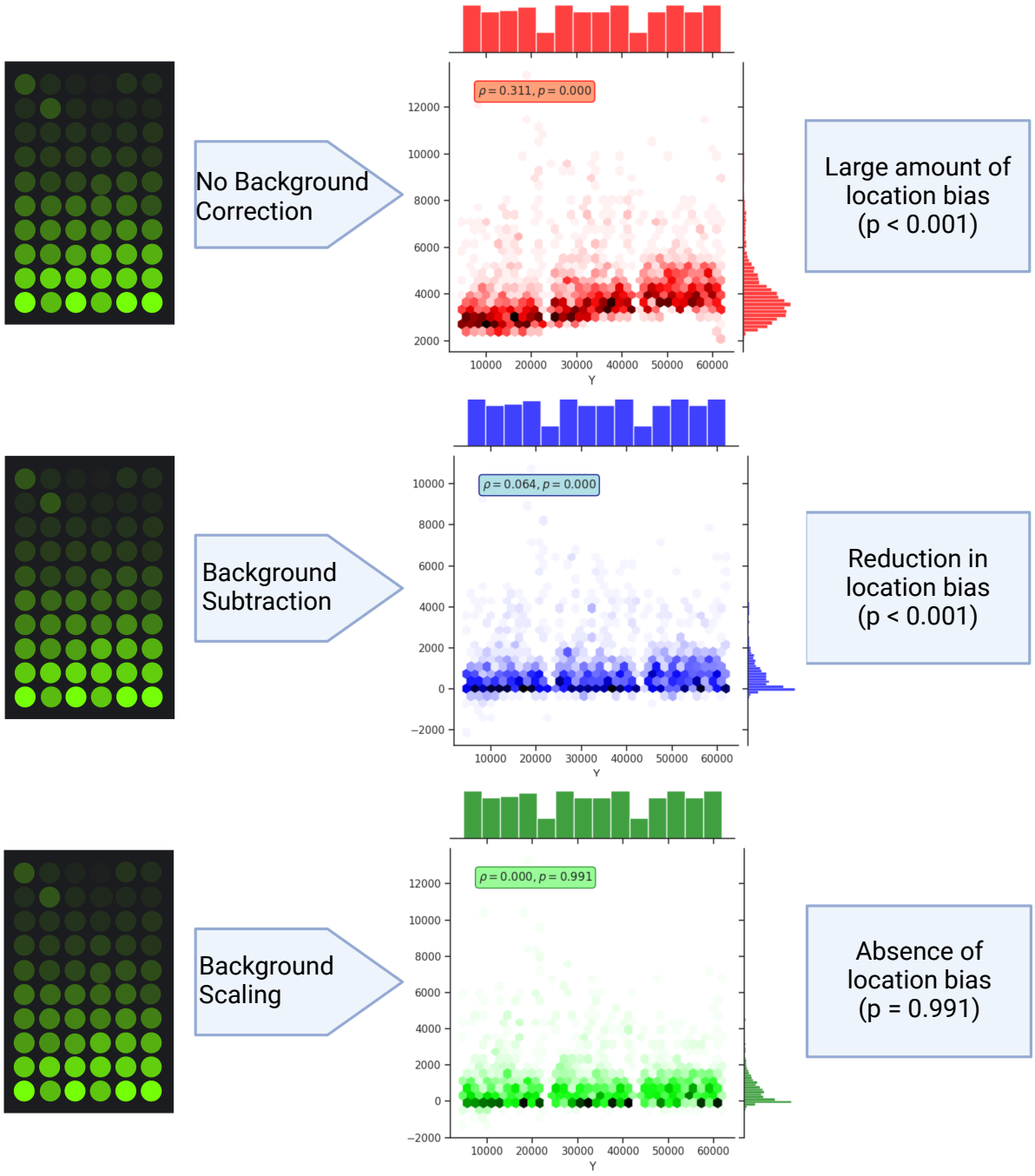

Figure 6.2: Demonstration of Location Bias. A. A simplified representation of an array (rotated 90 degrees), with a histogram demonstrating intensity values above it. In this histogram, the x-axis is fluorescence intensity and the colour indicates the density of histogram, with darker colour indicating the presence of more data points. As a result of the location bias, it is impossible to interpret this array without application of a suitable background correction measure. **B.** An array that has had the location bias removed by correcting for the background. This array facilitates easy interpretation of the results and determination of the high-intensity probes.

Array		Raw		Subtracted		Scaled		Scaling > Subtraction
	Treatment	Spearman ρ	p value	Spearman ρ	p value	Spearman ρ	p value	
S1 Day 1	Control	0.311	0.000	0.064	0.000	0.000	0.991	TRUE
S1 Day 2	Control	0.110	0.000	0.025	0.171	-0.001	0.943	TRUE
S1 Day 3	Control	0.174	0.000	0.034	0.061	0.013	0.480	TRUE
S2 Day 1	Control	0.234	0.000	0.023	0.205	-0.032	0.079	FALSE
S2 Day 2	Control	0.106	0.000	0.059	0.001	0.047	0.009	TRUE
S2 Day 3	Control	0.043	0.018	0.056	0.002	0.061	0.001	FALSE
S3 Day 1	Control	-0.108	0.000	-0.026	0.148	-0.003	0.883	TRUE
S3 Day 2	Control	0.335	0.000	0.111	0.000	-0.001	0.945	TRUE
S3 Day 3	Control	-0.071	0.000	0.071	0.000	0.039	0.033	TRUE
S4 Day 1	LPS-Cortisol	0.242	0.000	-0.034	0.058	0.028	0.128	TRUE
S4 Day 2	LPS-Cortisol	0.195	0.000	0.105	0.000	0.073	0.000	TRUE
S4 Day 3	LPS-Cortisol	0.209	0.000	0.086	0.000	0.042	0.022	TRUE
S5 Day 1	LPS-Cortisol	0.157	0.000	0.024	0.183	-0.007	0.691	TRUE
S5 Day 2	LPS-Cortisol	0.278	0.000	0.126	0.000	0.065	0.000	TRUE
S5 Day 3	LPS-Cortisol	0.171	0.000	0.030	0.095	-0.011	0.556	TRUE
S6 Day 1	LPS-Cortisol	0.370	0.000	0.033	0.069	-0.032	0.780	TRUE
S6 Day 2	LPS-Cortisol	0.331	0.000	0.051	0.005	-0.008	0.658	TRUE
S6 Day 3	LPS-Cortisol	0.104	0.000	0.082	0.000	-0.021	0.241	TRUE
S7 Day 1	LPS	0.326	0.000	0.075	0.000	-0.009	0.619	TRUE
S7 Day 2	LPS	0.180	0.000	0.060	0.001	0.037	0.042	TRUE
S7 Day 3	LPS	0.090	0.000	0.028	0.119	0.005	0.793	TRUE
S8 Day 1	LPS	0.259	0.000	0.028	0.118	-0.021	0.246	TRUE
S8 Day 2	LPS	0.239	0.000	0.072	0.000	0.024	0.179	TRUE
S8 Day 3	LPS	0.221	0.000	0.055	0.003	-0.015	0.400	TRUE
S9 Day 1	LPS	0.362	0.000	0.062	0.001	0.024	0.184	TRUE
S9 Day 2	LPS	0.328	0.000	0.059	0.001	0.014	0.438	TRUE
S9 Day 3	LPS	0.270	0.000	0.095	0.000	0.025	0.173	TRUE
Mean		0.216	0.000	0.058	0.003	0.024	0.117	0.926

Table 6.1: Comparison between Uncorrected, Background Subtracted, and Background Scaled Datasets using data from the phosphorylation-specific stain subset of the LPS-stimulation dataset. This table contains five major columns, and the first four columns have two subcolumns. The "Array" column provides descriptive information regarding the experimental layout in the descriptor of the array and the treatment group. The next three columns each contain a Spearman ρ value and a p value. The "Raw" column has no pre-processing applied, using the uncorrected foreground values, the "Subtracted" column uses the typical background subtraction method, and the "Scaled" column uses the new background scaling technique. The final column is TRUE if background scaling outperforms background subtraction, and FALSE otherwise. "Spearman ρ" is calculated by scipy.stats.spearmanr as the correlation between probe location and probe intensity, and the column mean is calculated as the arithmetic mean of all values in the column. "p value" is calculated by scipy.stats.spearmanr as the probability of an uncorrelated system producing a dataset with an equal Spearman ρ value to that of the array, and the column mean is calculated as the geometric mean. For the final column, the value in the "Mean" row is the fraction of arrays in which background scaling outperformed background subtraction.

Array		Raw		Subtracted		Scaled		Scaling > Subtraction
	Treatment	Spearman ρ	p value	Spearman ρ	p value	Spearman ρ	p value	
ATP1 Day 1	LPS	0.092	0.000	0.096	0.000	0.086	0.000	TRUE
ATP1 Day 2	LPS	-0.054	0.003	0.036	0.056	0.018	0.315	TRUE
ATP1 Day 3	LPS	0.268	0.000	0.133	0.000	0.029	0.105	TRUE
ATP2 Day 1	LPS	-0.147	0.000	0.016	0.376	-0.006	0.751	TRUE
ATP2 Day 2	LPS	0.176	0.000	0.152	0.000	0.024	0.141	TRUE
ATP2 Day 3	LPS	0.025	0.167	0.033	0.066	0.033	0.067	TRUE
ATP3 Day 1	LPS	0.171	0.000	0.158	0.000	0.086	0.000	TRUE
ATP3 Day 2	LPS	0.046	0.011	0.087	0.000	0.076	0.000	TRUE
ATP3 Day 3	LPS	0.145	0.000	0.089	0.000	0.013	0.461	TRUE
ATP4 Day 1	LPS-Cortisol	0.217	0.000	0.158	0.000	0.072	0.000	TRUE
ATP4 Day 2	LPS-Cortisol	-0.058	0.002	0.020	0.261	-0.005	0.768	TRUE
ATP4 Day 3	LPS-Cortisol	0.151	0.000	0.278	0.000	0.040	0.026	TRUE
ATP5 Day 1	LPS-Cortisol	0.009	0.635	0.196	0.000	0.061	0.001	TRUE
ATP5 Day 2	LPS-Cortisol	-0.276	0.000	-0.106	0.000	0.017	0.354	TRUE
ATP5 Day 3	LPS-Cortisol	-0.096	0.000	-0.040	0.027	-0.005	0.789	TRUE
ATP6 Day 1	LPS-Cortisol	0.183	0.000	0.148	0.000	0.113	0.000	TRUE
ATP6 Day 2	LPS-Cortisol	0.367	0.000	0.176	0.000	0.081	0.000	TRUE
ATP6 Day 3	LPS-Cortisol	0.193	0.000	0.026	0.152	-0.097	0.000	FALSE
ATP7 Day 1	Control	0.173	0.000	0.178	0.000	0.096	0.000	TRUE
ATP7 Day 2	Control	-0.078	0.000	-0.025	0.183	0.011	0.553	TRUE
ATP7 Day 3	Control	-0.008	0.659	-0.031	0.092	-0.018	0.322	TRUE
ATP8 Day 1	Control	0.084	0.000	0.074	0.000	0.067	0.000	TRUE
ATP8 Day 2	Control	0.266	0.000	0.190	0.000	0.063	0.002	TRUE
ATP8 Day 3	Control	0.082	0.000	0.013	0.468	-0.049	0.008	FALSE
ATP9 Day 1	Control	0.094	0.000	0.075	0.000	0.064	0.000	TRUE
ATP9 Day 2	Control	-0.263	0.000	0.237	0.000	0.090	0.000	TRUE
ATP9 Day 3	Control	0.091	0.000	0.018	0.321	-0.055	0.003	TRUE
Mean		0.141	0.000	0.104	0.001	0.051	0.005	0.926

Table 6.2: Comparison Between No Correction, Background Subtraction, and Background Scaling using data from the fluorescent-labelled ATP subset of the LPS-stimulation dataset. This table contains five major columns, and the first four columns have two subcolumns. The "Array" column provides descriptive information regarding the experimental layout, i.e., the descriptor of the array and the treatment group. The next three columns each contain a Spearman ρ value and a p value. The "Raw" column has no pre-processing applied, using the uncorrected foreground values, the "Subtracted" column uses the typical background subtraction method, and the "Scaled" column uses the new background scaling technique. The final column is TRUE if background scaling outperforms background subtraction, and FALSE otherwise. "Spearman ρ" is calculated by scipy.stats.spearmanr as the correlation between probe location and probe intensity, and the column mean is calculated as the arithmetic mean of all values in the column. "p value" is calculated by scipy.stats.spearmanr as the probability of an uncorrelated system producing a dataset with an equal Spearman ρ value to that of the array, and the column mean is calculated as the geometric mean. For the final column, the value in the "Mean" row is the fraction of arrays in which background scaling outperformed background subtraction.

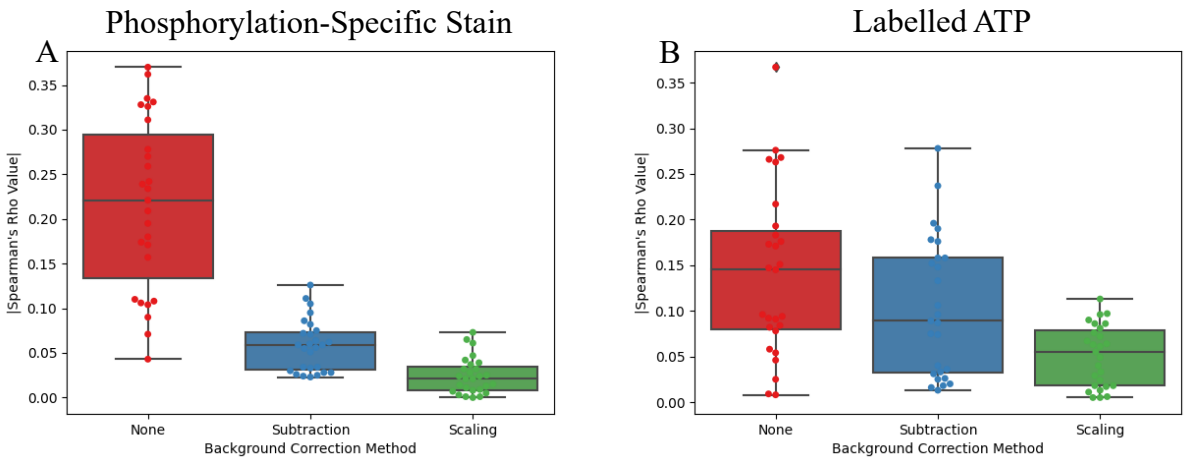

Figure 6.3: Summary of the Effect of Applying Background Scaling to the LPS Stimulation dataset. **A.** Comparison between background subtraction, background scaling, and no background correction for all 27 arrays of the phosphorylation-specific stain subset of the LPS stimulation dataset. Points represent individual arrays, colour-coded by treatment (blue, LPS; red, LPS–cortisol; green, control). Whiskers and box edges of the box plots represent data quartiles. Center line represents the median value. The y-axis is the absolute value Spearman's ρ value for the correlation between location on the array and probe intensity. It can be seen than the background scaling method decreases the location–intensity correlation. **B.** As in A, but for the fluorescent-labelled ATP subset of the LPS stimulation dataset. Again, it can be seen than the background scaling method decreases the location–intensity correlation.

From the above figures and tables, it can be observed that the background scaling method outperforms background subtraction in correction of location bias. For the phosphorylation-specific stain, background scaling had a greater reduction in location bias than background subtraction in 25 out of 27 arrays (92.6%), as quantified by the value of Spearman's rank correlation coefficient (Spearman's ρ) for the correlation between intensity value and physical position on the array along the y-axis. No significant (p <0.05) location–intensity correlation was found in 20 arrays (74.1%) using background scaling. In comparison, background subtraction resulted in no significant location–intensity correlation in just 9 arrays (33.3%). The mean Spearman ρ values of the uncorrected, background subtracted, and background scaled arrays were 0.216, 0.058, and 0.024, respectively, suggesting a progressive and large reduction in location bias across the methods. In addition, the background scaling method is the only one of the three to have a geometric mean of the p values above the significance threshold. For the fluorescent-labelled ATP subset, background scaling provided a superior reduction in location bias than background subtraction in 25 out of 27 arrays (92.6%). No significant (p <0.05) location–intensity correlation was found in 11 arrays (40.7%) using background scaling, while when using background subtraction, no significant location–intensity correlation was found in 9 arrays (33.3%). The mean Spearman ρ values of the uncorrected, background subtracted, and background scaled arrays were 0.141, 0.104, and 0.51, respectively, clearly showing the superiority of the background scaling method in this experiment.

Figure 6.4 provides a simplified overview of the purported benefit of background subtraction.

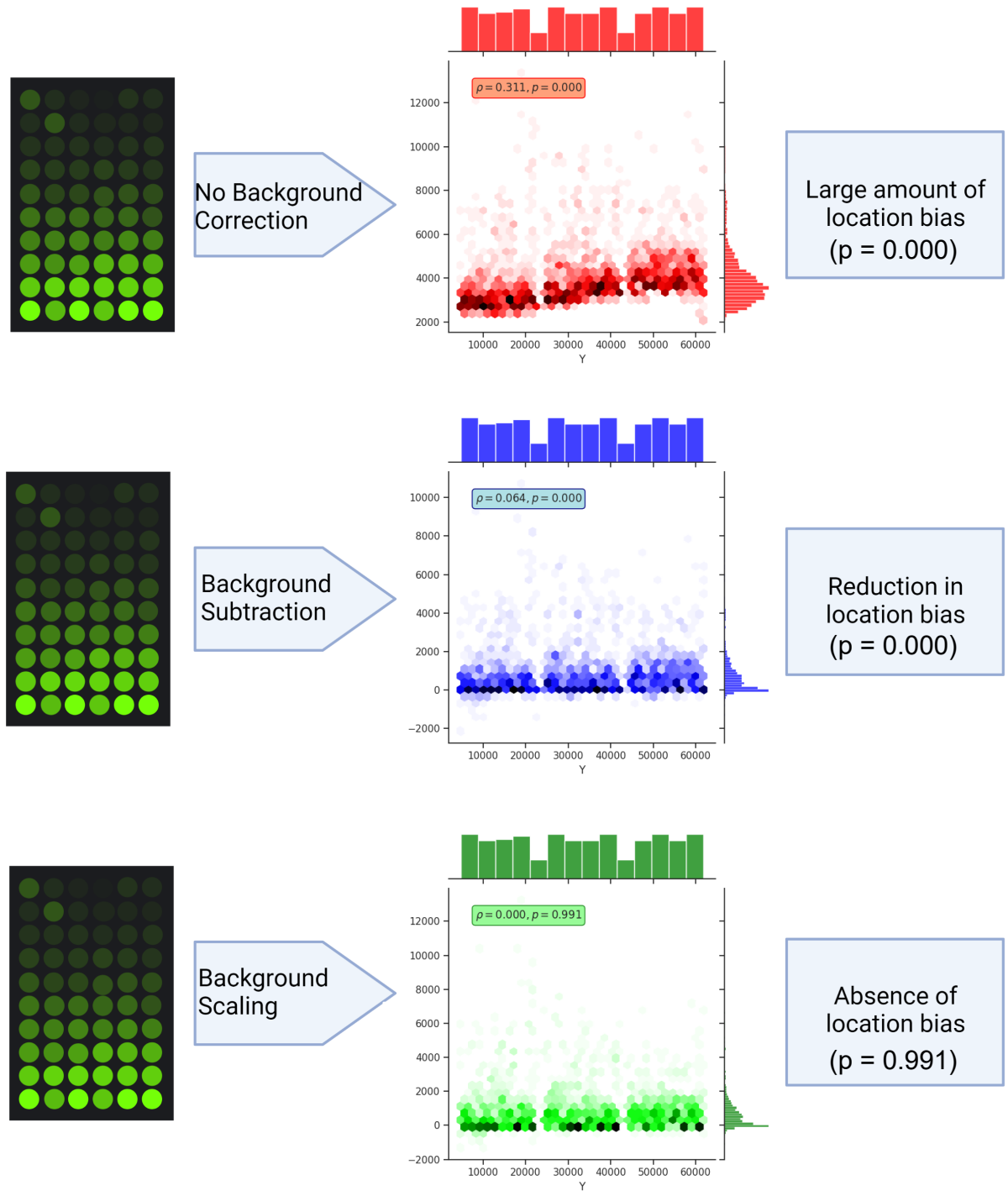

Figure 6.4: Simple Summary of Location Bias Correction. Data displayed in the center plots represent that from the phosphorylation-specific stain subset of the LPS stimulation dataset, day 1, sample 1, cortisol–LPS treatment group. X-axes and frequency histograms along the x-axes represent the position of spots on the array (identical for all three histograms). Y-axes and frequency histograms along the y-axes represent probe intensities, which concentrate at lower values when correction is applied. Each panel has a Spearman ρ value, calculated by scipy.stats.spearmanr as the correlation between probe location and probe intensity and a p value, calculated by scipy.stats.spearmanr as the probability of an uncorrelated system producing a dataset with an equal Spearman ρ value to that of the array, and the column mean is calculated as the geometric mean.

6.3 Application of PIIKA2.5 on the LPS Stimulation Dataset

This analysis concerns the investigation of the LPS stimulation dataset. As a case study, this dataset provides valuable insight. Its design is based on a well-defined interaction: LPS stimulation of PBMCs is known to induce toll-like receptor signalling that manifests in the induced release of pro-inflammatory cytokines such as TNFa [78]. The pre-existing knowledge of the experimental outcome is important for comparison of the results before and after the application of PIIKA 2.5. A detailed description of this dataset and the experimental procedure can be found in Section 6.1.1. Spot size analysis is not included in this section as the arrays from this experiment were properly manually realigned prior to the analysis described here.

6.3.1 Background Scaling

A common source of systematic error in kinome microarrays is variation in intensity (background and foreground) in relation to the spatial characteristics of an array (Section 2.3.1). While background subtraction alone is effective at reducing such location biases, the background scaling calculation described in Section(4.1) and validated in Section 6.2 further decreases the location-associated variation. Background scaling is similar to background subtraction except that it involves an additional factor that relates local background to the background across the entire array.

The reduction in location bias corrects for inherent location biases present in kinome microarrays to a degree greater than background subtraction (Figure 6.5). Of the 27 arrays analysed, 74.07% of arrays showed a reduction of the Pearson correlation coefficient between the Y-axis position of a probe and the probe foreground intensity, with the average significance of the correlation decreasing from 0.07 to 0.32 (Table 6.1). Without correction, this bias is present (p <0.0001, based on the significance of the Pearson correlation coefficient) in all 27 arrays analysed in the dataset. A binomial test performed on the data reveals a significant effect on the probability of reducing the location bias ($p = 0.0066$) by using background scaling over background subtraction, though the small sample size of 27 should be taken into consideration when interpreting this result.

6.3.2 Outlier Detection

Arrays that are significant outliers indicate a systematic error that can have consequences on data analysis. This error is not often correlated with intended experimental factors and appears to be independent of treatment. Within the twenty-seven arrays from the LPS stimulation dataset under consideration, two arrays, occurring on different experimental days and reflecting different treatment conditions, were flagged as being problematic (Figure 6.6). For each array, a distance is calculated by summing the mean absolute differences between corresponding probes. Arrays with a significantly high sum of distances (p <0.05) as determined by one-way ANOVA are flagged as outliers. For each array, the distance between it and all other arrays in the experiment can be summed. By determining the sum corresponding to a p value of 0.05 from

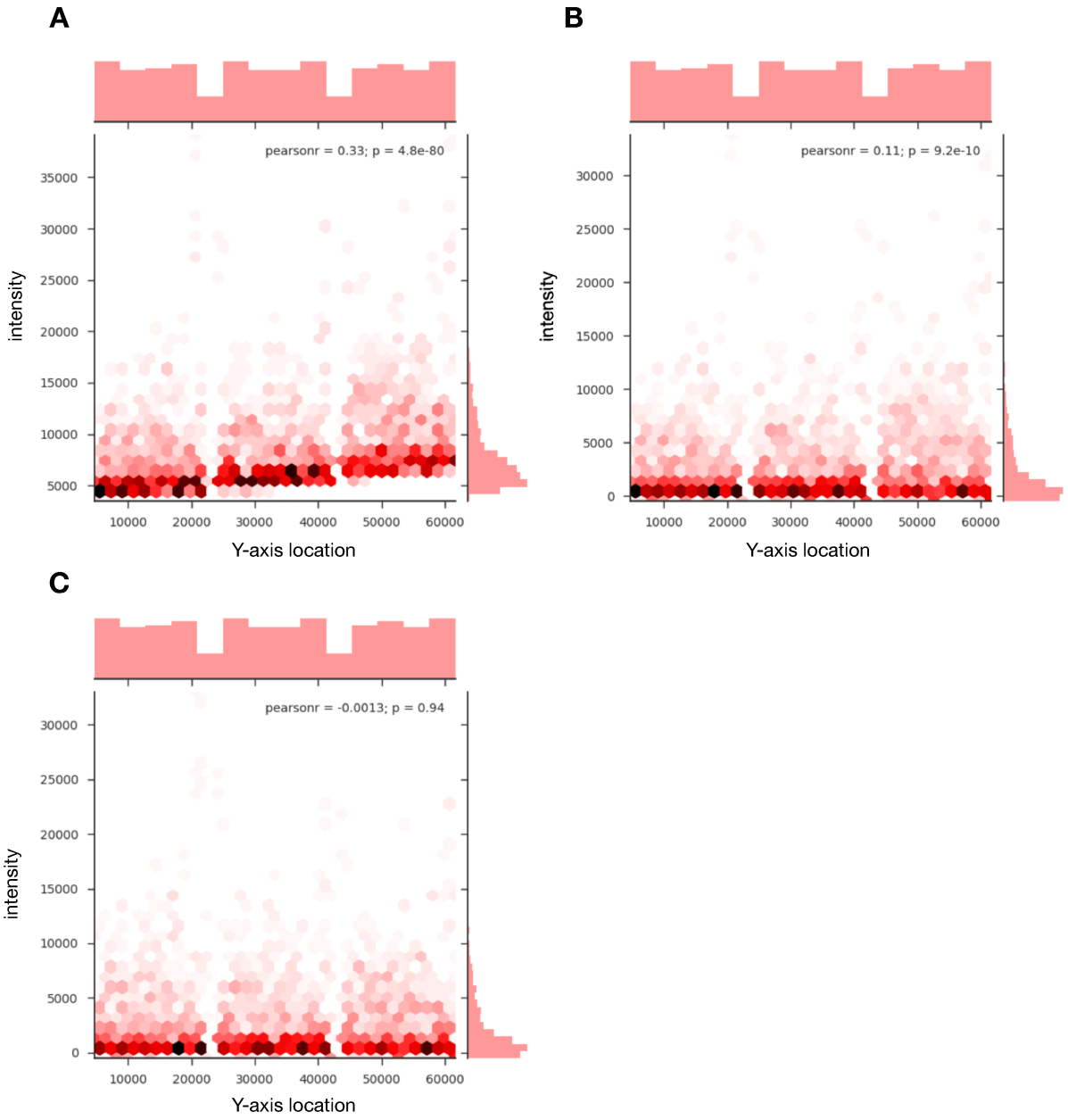

Figure 6.5: Background Scaling to Account for Regional Variations A. Uncorrected output from an array showing a high degree of bias, creating a correlation between probe foreground intensity and the spot location on the y-axis of the array. B. Output from the same array as A, corrected using background subtraction. The bias has reduced substantially, but there still is a discernible and statistically significant correlation, especially in the probes with higher intensities. C. Output from the same array as A and B, corrected using background scaling, completely removing the correlation between probe intensity and y-axis. Data represents an array taken from the LPS stimulation experiment described earlier that is neither an outlier nor required any manual alignment correction (mean–median slope >0.80). Frequency histograms along the x-axis show the distribution of the position of spots on the array (identical for all three panels). Frequency histograms along the y-axis show the distribution of intensities which concentrate at lower values when correction is applied.

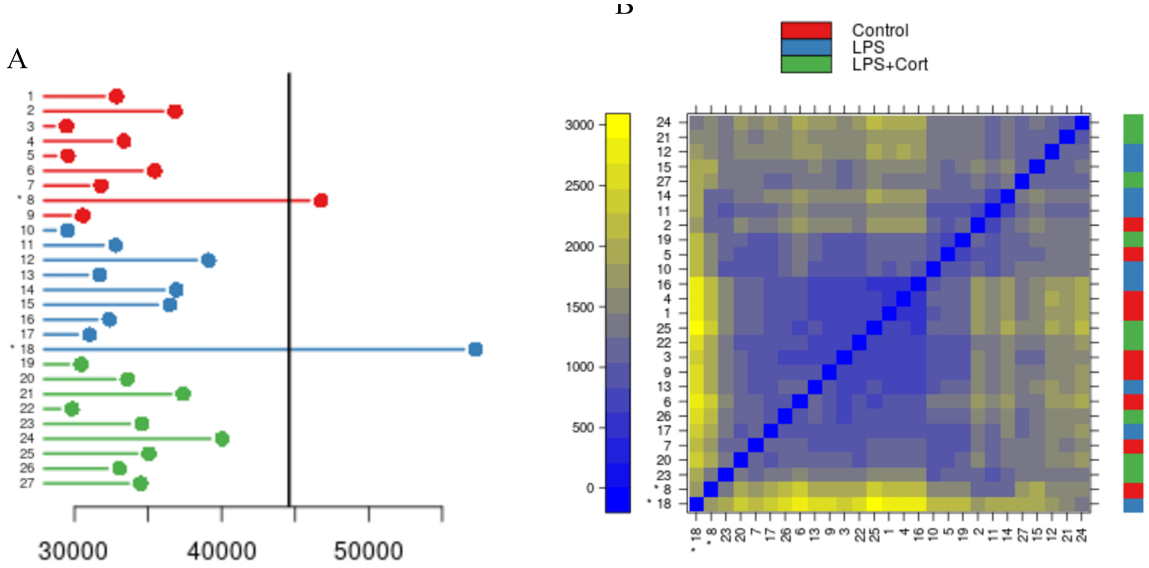

Figure 6.6: Identification of Outliers by Inter-Array Comparison. **A.** Sum of pairwise distances between the given array and all other arrays (N = 27, consisting of 3 biological replicates with 3 technical replicates each per treatment). The indicator line represents the point where the sum for a given array is significantly greater than the others with a p value of <0.05 according to one-way ANOVA with Tukey's HSD test, corrected for multiple comparisons. Datasets 8 and 18 are outliers with p values of 0.021 and <0.001, respectively. **B.** Heatmap showing pairwise distances between arrays. These are sorted from least significant to most significant difference of sums on the y-axis and the reverse on the x-axis. The color bar on the right indicates the treatment group of the array shown on the x-axis, with red representing control, blue representing LPS alone, and green representing LPS with cortisol. Datasets 8 and 18 are present in the bottom left of the heatmap and are marked by asterisks.

the ANOVA test, a quick visual synopsis of the variation in datasets is possible (Figure 6.6A). A "stoplight" system is used to convey the results to the user, assigning either green, amber, or red to each array, with green indicating that there is no apparent issue with the array, red indicating that the array is an outlier, and amber suggesting caution in subsequent analyses. For the dataset from the phosphorylation-specific stain subset of the LPS stimulation experiment, there are no arrays flagged as amber ($0.05 < p < 0.1$), but arrays 8 and 18 ($p < 0.05$) receive a red light and are denoted as outliers. All other arrays receive a green light.

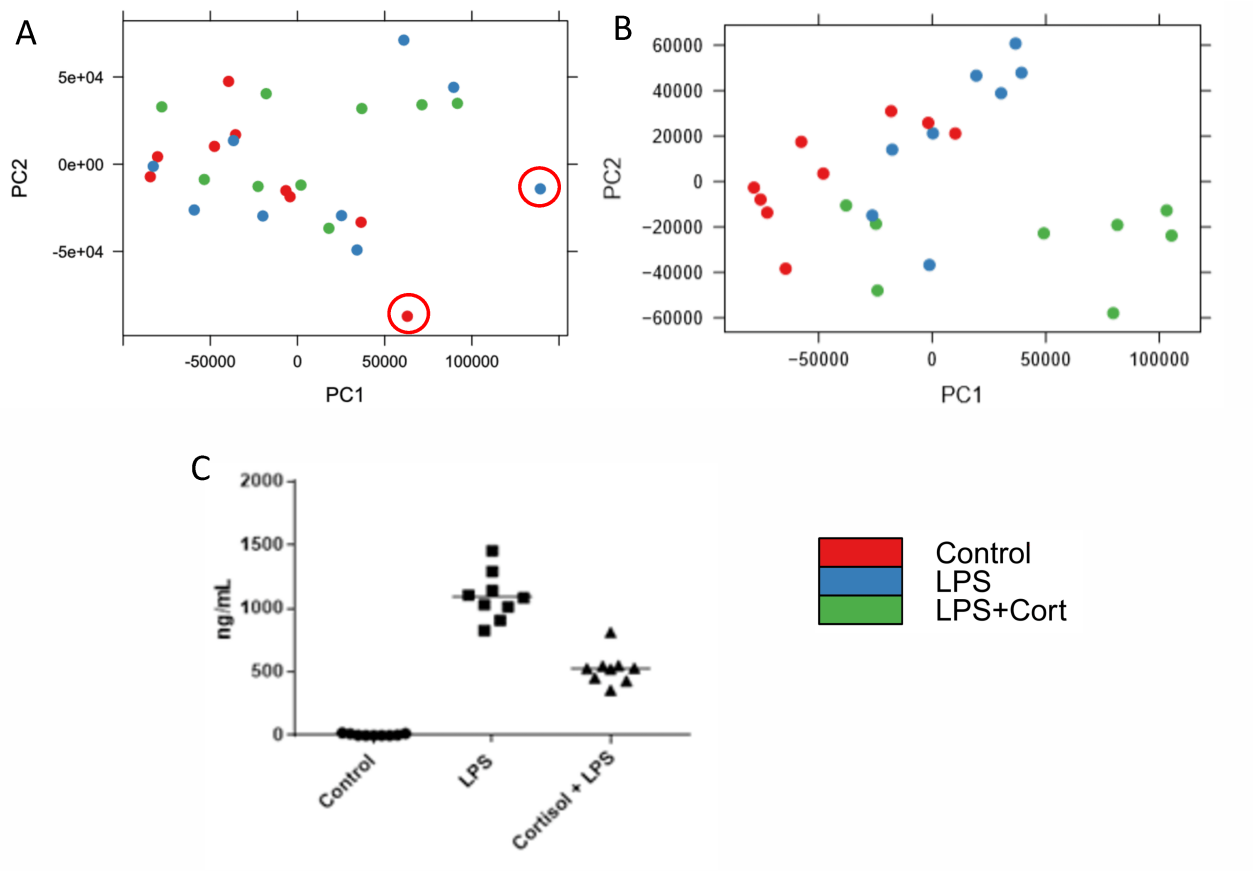

Figure 6.7: Principal Component Analysis A. Before and **B.** After Removal of the Arrays Determined to be Outliers. The two outliers removed are highlighted in panel A. This data represents 2 treatment groups (blue and green) and one control (red) with 3 biological replicates performed on three different days. **C.** ELISA data showing differential TNFa levels across the different treatment groups. This shows a strong difference in the immune response, which leads to the expectation that the PCA should show distinct clustering.

The presence of two outliers has important consequences to data analysis. For example, principal component analysis performed with the inclusion of the red-flagged datasets fails to separate the datasets on the basis of treatment condition (Figure 6.7A). The same analysis, performed after removal of the two flagged datasets, offers improved separation of the datasets according to treatment groups (Figure 6.7B). This was confirmed by calculating both the Dunn index and Davis-Bouldin Index (DBI) before and after removal of the outlier arrays. Before outlier removal, the Dunn index and DBI were 0.62 and 0.82, respectively. Following removal, these values were improved to 0.18 (DBI) and 0.97 (Dunn index).

Cortisol serves to dampen pro-inflammatory responses, including those induced by TLR agonists such as LPS. The pre-treatment of the cells with cortisol results in a clear and consistent reduction of LPS-induced TNFa release (6.7C).

6.4 Application of PIIKA2.5 on the Breast Cancer Dataset

The following section details the application of PIIKA 2.5's quality control features to the breast cancer dataset described in Section6.1.2. Prior to the work described herein, these data have been difficult to analyse due to large differences in intensity values between the two days that the experiment was performed over. In other words, the results from these two days are more similar between arrays processed on the same day than they are on the basis of treatment group (healthy vs. cancer vs. benign). An equal number of samples on identically designed microarrays were processed using the same procedure on two different days. The results showed a more significant effect of the day of the experiment than of the samples' treatment group. Given this context, rectification of this dataset would be very strong evidence of the effectiveness of PIIKA 2.5.

6.4.1 Background Scaling

Background scaling was applied to the breast cancer dataset prior to all other downstream analyses. Days were assessed separately due to the aforementioned experimental variability between the two days. A comparison of the background correction techniques (no correction, background subtraction, and background scaling) can be seen in Table 6.3 for day 1 and Table 6.4 for day 2.

In the context of the breast cancer data, it can be observed that the background scaling method outperformed background subtraction in the correction of location bias. For the data from day 1, background scaling had a greater reduction in location bias than background subtraction in 23 out of 30 arrays (76.7%). No significant (p <0.05) location–intensity correlation was found in 13 arrays (40.7%) using background scaling, while when using background subtraction, no significant location–intensity correlation was found in 3 arrays (10.0%). For the data from day 2, background scaling had a greater reduction in location bias than background subtraction in 22 out of 30 arrays (73.3%). Identically to day 1, 13 arrays (40.7%)had no significant location–intensity correlation when using background scaling and 3 arrays (10.0%) when using background subtraction. It is an interesting note that background scaling, while still outperforming the other two methods, performed worse here than on the experiments presented in Section 6.2, both overall and in comparison to background subtraction. The most obvious discrepancy between these datasets, and therefore the most likely culprit of the difference in performance, is that the breast cancer dataset has a much stronger location bias than the LPS stimulation dataset, as evidenced by the Spearman ρ values of the uncorrected data (0.375 and 0.431 for breast cancer days 1 and 2 vs. 0.216 and 0.141 for the phosphorylation-specific stain and the fluorescent-labelled ATP).

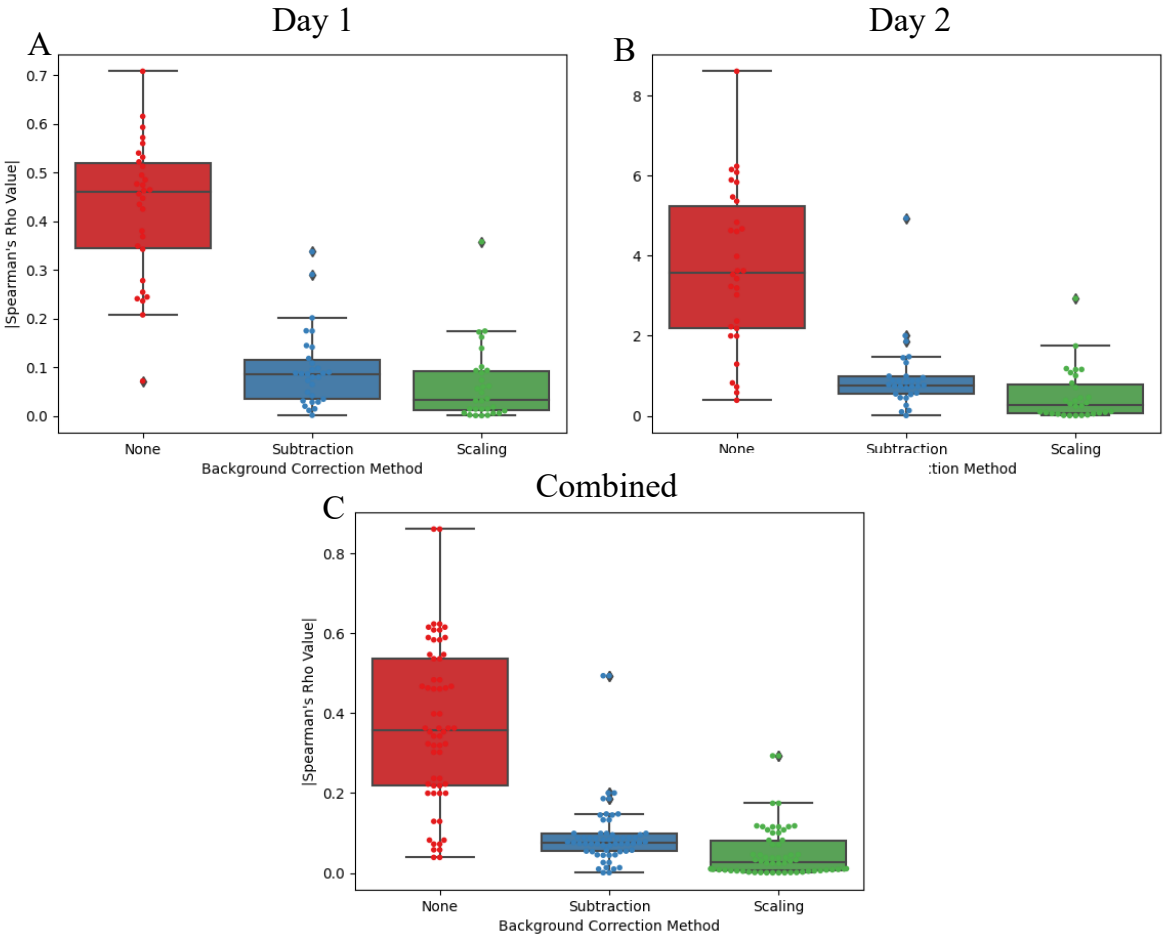

Figure 6.8: Summary of Background Correction Methods Applied to The Breast Cancer Dataset. A. Comparison between background subtraction, background scaling, and no background correction for all 30 arrays of the breast cancer dataset on day 1. Points represent individual arrays, colour-coded by sample type (blue, benign; red, cancer; green, healthy). Whiskers and box edges of the box plots represent data quartiles, and the interior line represents the median value. The y-axis is the absolute value of Spearman's ρ for the correlation between probe location and probe intensity. It can be seen that the background scaling method decreases the location–intensity correlation. **B.** As in A, but for the arrays from day 2 of the same experiment. Again, it can be seen than the background scaling method decreases the location–intensity correlation. **C.** Comparison between background subtraction, background scaling, and no background correction for all 60 arrays of the entire breast cancer dataset. Points represent individual arrays, colour-coded by experiment day (blue, day 1; red, day 2).

Array	Raw		Subtracted		Scaled		Scaling > Subtraction
	Spearman ρ	p value	Spearman ρ	p value	Spearman ρ	p value	
B1	0.425	0.000	0.034	0.000	0.061	0.000	FALSE
B2	0.244	0.000	0.014	0.102	-0.031	0.143	FALSE
B3	0.495	0.000	0.291	0.000	0.002	0.839	TRUE
B4	0.475	0.000	0.174	0.000	0.172	0.000	TRUE
B5	0.522	0.000	-0.041	0.000	-0.014	0.110	TRUE
B6	0.477	0.000	0.175	0.000	0.162	0.000	TRUE
B7	0.540	0.000	0.108	0.000	0.006	0.497	TRUE
B8	0.465	0.000	0.049	0.000	0.059	0.000	FALSE
B9	0.560	0.000	-0.080	0.000	-0.035	0.000	TRUE
B10	0.380	0.000	0.073	0.000	-0.001	0.893	TRUE
C1	-0.236	0.000	0.145	0.000	0.014	0.112	TRUE
C2	0.254	0.000	-0.020	0.024	0.014	0.120	TRUE
C3	0.241	0.000	-0.001	0.886	0.000	0.962	TRUE
C4	0.513	0.000	0.028	0.001	0.010	0.238	TRUE
C5	0.463	0.000	0.011	0.203	0.005	0.542	TRUE
C6	0.208	0.000	0.065	0.000	0.056	0.000	TRUE
C7	0.532	0.000	0.087	0.000	0.093	0.000	FALSE
C8	0.485	0.000	-0.028	0.001	0.014	0.105	TRUE
C9	0.708	0.000	0.337	0.000	0.358	0.000	FALSE
C10	0.343	0.000	0.098	0.000	0.093	0.000	TRUE
H1	0.072	0.000	0.031	0.000	0.031	0.000	FALSE
H2	0.616	0.000	0.087	0.000	0.139	0.000	FALSE
H3	0.447	0.000	0.086	0.000	0.073	0.000	TRUE
H4	0.349	0.000	0.080	0.000	-0.001	0.941	TRUE
H5	0.368	0.000	0.088	0.000	-0.005	0.532	TRUE
H6	0.593	0.000	0.141	0.000	0.101	0.000	TRUE
H7	0.572	0.000	0.095	0.000	-0.018	0.038	TRUE
H8	0.435	0.000	0.118	0.000	0.089	0.000	TRUE
H9	-0.278	0.000	0.201	0.000	0.174	0.000	TRUE
H10/P	0.456	0.000	0.089	0.000	0.048	0.000	TRUE
Mean	0.431	0.000	0.094	0.000	0.064	0.004	0.852

Table 6.3: Comparison Between Background Correction Methods Applied to the Breast Cancer Dataset (Day 1). In the "Array" column, the letter prior to the sample number indicates the health status of the sample. H signifies a healthy sample, B signifies benign, and C signifies cancer. P signifies papilloma; since there is only one papilloma sample, it is grouped with the healthy samples for statistical analyses. The next three columns each contain a Spearman ρ value and a p value. The "Raw" column has no pre-processing applied, using the uncorrected foreground values, the "Subtracted" column uses the typical background subtraction method, and the "Scaled" column uses the new background scaling technique. The final column is TRUE if background scaling outperforms background subtraction, and FALSE otherwise. "Spearman ρ" is calculated with the python function scipy.stats.spearmanr as the correlation between probe location and probe intensity, and the column mean is calculated as the arithmetic mean of all values in the column. "p value" is calculated with scipy.stats.spearmanr as the probability of an uncorrelated system producing a dataset with a Spearman ρ value equal to that of the array, and the column mean is calculated as the geometric mean. For the final column, the value in the "Mean" row is the fraction of arrays in which background scaling outperformed background subtraction.

Array	Raw		Subtracted		Scaled		Scaling > Subtraction
	Spearman ρ	p value	Spearman ρ	p value	Spearman ρ	p value	
B1	0.584	0.000	-0.073	0.000	-0.033	0.000	TRUE
B2	0.058	0.000	0.099	0.000	0.100	0.000	FALSE
B3	0.461	0.000	-0.095	0.000	-0.045	0.000	TRUE
B4	-0.218	0.000	-0.053	0.000	-0.071	0.000	FALSE
B5	0.615	0.000	-0.045	0.000	0.011	0.216	TRUE
B6	0.082	0.000	0.074	0.000	0.082	0.000	FALSE
B7	0.484	0.000	0.026	0.003	-0.002	0.790	TRUE
B8	0.343	0.000	0.073	0.000	0.000	0.963	TRUE
B9	0.399	0.000	0.200	0.000	-0.005	0.554	TRUE
B10	0.223	0.000	0.001	0.928	0.030	0.001	FALSE
C1	0.467	0.000	-0.088	0.000	-0.046	0.000	TRUE
C2	0.199	0.000	-0.074	0.000	-0.038	0.000	TRUE
C3	-0.072	0.000	0.057	0.000	-0.010	0.258	TRUE
C4	0.323	0.000	0.087	0.000	0.115	0.000	FALSE
C5	0.319	0.000	0.010	0.262	-0.001	0.953	TRUE
C6	0.237	0.000	0.076	0.000	0.001	0.907	TRUE
C7	0.463	0.000	-0.078	0.000	-0.025	0.005	TRUE
C8	0.039	0.000	0.186	0.000	0.175	0.000	TRUE
C9	-0.129	0.000	0.133	0.000	0.118	0.000	TRUE
C10	0.362	0.000	-0.044	0.000	-0.001	0.891	TRUE
H1	0.547	0.000	-0.013	0.131	0.024	0.007	FALSE
H2	0.589	0.000	-0.069	0.000	-0.004	0.611	TRUE
H3	0.608	0.000	-0.099	0.000	-0.039	0.000	TRUE
H4	0.362	0.000	-0.084	0.000	0.006	0.514	TRUE
H5	0.861	0.000	-0.494	0.000	-0.293	0.000	TRUE
H6	0.354	0.000	-0.055	0.000	-0.009	0.310	TRUE
H7	0.623	0.000	-0.088	0.000	-0.008	0.363	TRUE
H8	-0.200	0.000	0.145	0.000	0.116	0.000	TRUE
H9	0.536	0.000	0.148	0.000	0.007	0.395	TRUE
H10	0.302	0.000	0.087	0.000	0.108	0.000	FALSE
	0.375	0.000	0.096	0.000	0.047	0.009	0.815

Table 6.4: Comparison Between Background Correction Methods Applied to the Breast Cancer Dataset (Day 2). In the "Array" column, the letter prior to the sample number indicates the health status of the sample. H signifies a healthy sample, B signifies benign, and C signifies cancer. The next three columns each contain a Spearman ρ value and a p value. The "Raw" column has no pre-processing applied, using the uncorrected foreground values, the "Subtracted" column uses the typical background subtraction method, and the "Scaled" column uses the new background scaling technique. The final column is TRUE if background scaling outperforms background subtraction, and FALSE otherwise. "Spearman ρ" is calculated with the python function scipy.stats.spearmanr as the correlation between probe location and probe intensity and the column mean is calculated as the arithmetic mean of all values in the column. "p value" is calculated with scipy.stats.spearmanr as the probability of an uncorrelated system producing a dataset with a Spearman *rho* value equal to that of the array and the column mean is calculated as the geometric mean. For the final column, the value in the "Mean" row is the fraction of arrays in which background scaling outperformed background subtraction.

6.4.2 Detection of Improper Spot Size

The breast cancer dataset has been previously manually corrected, and has no arrays with a slope of the background-corrected mean to the background-corrected median below either flagging threshold (amber or red light threshold). The lowest mean–median slope of all of the arrays in the breast cancer dataset is 0.832 (Sample number 110 from day 1), above the threshold required for an array to be given an amber light. The mean–median linear regression plot of this array is shown in Figure 6.9 as an example.

The average background-corrected mean–median slope of the breast cancer dataset is 0.9167, with no significant difference between day 1 (0.9142) and day 2 (0.919) (Figure 6.10). However, there is an apparent difference in the variance of the mean–median slopes between the days, as evidence by difference in the quartile sizes of day 1 and day 2 in (Figure 6.10). Regardless, the variation is not enough to cause any array to cross the threshold required for it to be flagged with either a red or amber light. The error caused by improper spot size has been nearly eliminated in the breast cancer dataset through manual correction. As a result of the high level of quality, no corrective action (manual resizing) was recommended by PIIKA 2.5 and no correction was needed.

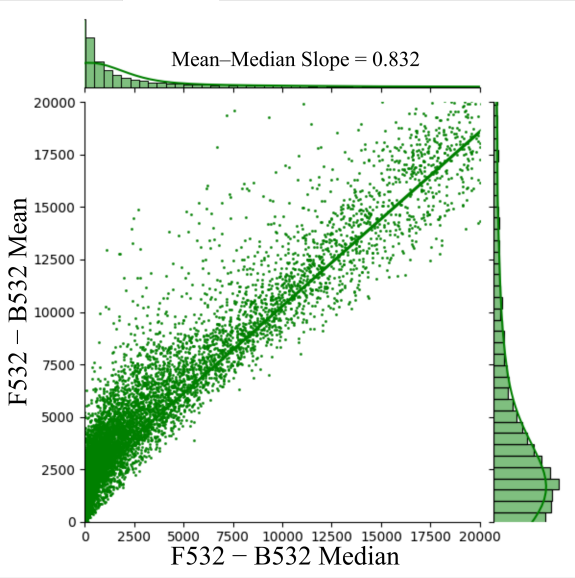

Figure 6.9: Example Mean-–Median Plot from the Breast Cancer Dataset. Individual points on the plot each represent a single probe on a given array. In this specific figure, data are a cancer sample from day 1 of the experiment and represent the lowest mean–median slope value of 0.832. This result is above the amber threshold for determination of improper spot size and no further correction is required.

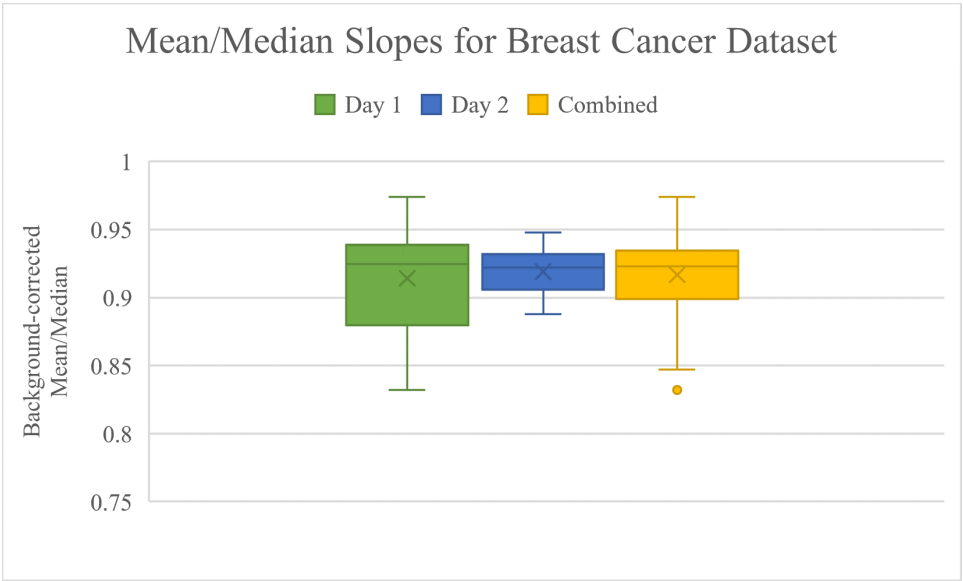

Figure 6.10: Summary of All Breast Cancer Mean—Median Slopes. The lowest line (0.75) represents PIIKA 2.5's amber threshold for determination of improper spot size. Day 1 mean mean–median slope: 0.9142; day 2 mean mean–median slope: 0.919; mean mean–median slope of both days combined: 0.9167

6.4.3 Outlier Detection

Lastly, outlier detection was applied to the breast cancer dataset. Within the breast cancer dataset, 10 samples were identified as outliers, all within the samples analysed on day 1 of the experiment with an equal distribution across the three major sample types (Table 6.5, Figure 6.11). As with previous analysis, there is no correlation between intentional experimental factors and the status of an array as an outlier, only between outlier status and day of the experiment (both p <0.001 as determined by Fisher's exact test). The singular papilloma sample was not an outlier and was grouped in the healthy samples so that each treatment group would have an equal number of samples. In addition to these nine outliers, there exists a borderline case, where the p value for outlier status is between 0.1 and 0.05, which would give this particular array an amber light from PIIKA 2.5.

The findings of both days analysed simultaneously are consistent with those of either day analysed individually, with day 1 having the same nine outliers and one borderline outlier, with day 2 having no outliers. This indicates that the outlier detection method used by PIIKA 2.5 is not sensitive the addition/removal of samples. This is also true in that removal of the outliers and re-evaluating the experiment does not create new outliers. In several other datasets it has been found that repeated outlier analysis can turn arrays bordering on significance into arrays flagged as outliers, but never will create outliers from clearly conforming arrays. This is specific to smaller datasets in which removal of an array is more substantial and more likely to affect the outlier analysis. Thus, in this case, once outliers are removed and outlier detection is repeated, the amber array (116, a benign sample) remains bordering on significance. It should be noted that repeating outlier analysis is highly discouraged for outlier detection [82].

Day 1			Day 2		
Sample #	Classification	Outlier	Sample #	Classification	Outlier
10	Benign	No	3	Healthy	No
12	Cancer	No	6	Benign	No
13	Healthy	No	8	Healthy	No
15	Cancer	Yes	14	Cancer	No
16	Benign	Yes	18	Cancer	No
26	Benign	No	21	Papilloma	No
29	Healthy	Yes	23	Benign	No
39	Cancer	No	32	Benign	No
43	Benign	Yes	35	Cancer	No
48	Healthy	No	37	Cancer	No
50	Benign	No	42	Healthy	No
52	Healthy	Yes	45	Healthy	No
60	Healthy	Yes	54	Healthy	No
72	Cancer	No	55	Healthy	No
74	Benign	Yes	58	Benign	No
79	Cancer	Yes	61	Cancer	No
82	Cancer	Yes	63	Benign	No
84	Healthy	No	66	Benign	No
87	Cancer	No	68	Benign	No
95	Cancer	No	71	Healthy	No
99	Benign	No	76	Healthy	No
105	Papilloma	No	81	Benign	No
108	Healthy	No	89	Healthy	No
110	Cancer	No	91	Cancer	No
112	Healthy	No	97	Benign	No
114	Benign	No	104	Cancer	No
116	Benign	No*	125	Cancer	No
120	Benign	No	128	Benign	No
123	Cancer	No	132	Cancer	No
130	Healthy	No	135	Benign	No

Table 6.5: Identification of Samples in the Breast Cancer Dataset * indicates an array bordering on significance threshold for outlier identification (p <0.1).

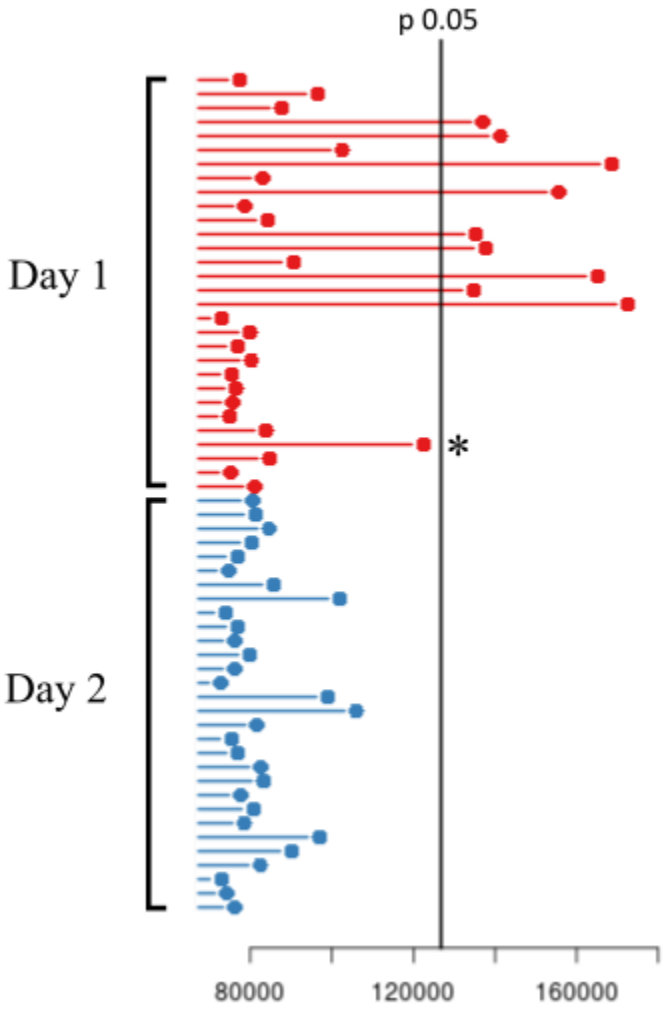

Figure 6.11: Outlier Identification in the Breast Cancer Dataset. Colours represent the day on which the experiment was performed. Each point represents the of pairwise distances between the given array and all other arrays (N = 60, consisting of 30 samples analysed on each of the two days of the experiment). The indicator line represents the point where the sum for a given array is significantly greater than the others with a p value of <0.05 according to one-way ANOVA with Tukey's HSD test, corrected for multiple comparisons using the Bonferroni correction. *, $p = 0.067$ (approaching significance). The distribution of outliers is completely uneven, with nine outliers on day one and zero outliers on day 2.

The lopsided distribution of outliers (nine on day 1, zero on day 2) is unique to this dataset and warranted investigation. GenePix's .gpr file output contains a large amount of data which is not utilized by PIIKA 2.5. A thorough investigation of this supplementary data did not reveal any factor or discernible trend present in the outlying arrays that was different between the outliers and the non-outliers, nor was there a discernible trend present exclusive to either experimental day. There are a number of theoretical conditions that the outlier distribution could be attributed to. For example, if an environmental factor like temperature or humidity of the external laboratory environment was different on one day of the experiment, it could cause a similar effect but this would be more likely to worsen temporal bias without creating an abundance of outliers. Indeed, it seems there are two separate effects present in this dataset: the regular temporal bias that can be seen in other multi-day experiments and some stronger effect that created the outliers on day 1. The source of this seems impossible to determine in post-experiment analyses. The presence of an experimental error such as contamination or improper sample storage cannot be ruled out.

In any case, the bias caused by the day of the experiment cannot be entirely accounted for by the fact that all of the outlier arrays were on the first of two experimental days. This is shown in Figure 6.12, a PCA visualization of the breast cancer dataset with both days included. It is clear that there is a significant clustering of the samples on day 1 (pictured in red), as well as of the samples from day 2 (pictured in blue). Additionally, it is not difficult to find which arrays are the outliers from observation of given PCA plot. A clear secondary group of arrays from day 1 are distinctly far away from the rest of the arrays. These are highlighted with a box, with a circle specifically highlighting array 27, the near-outlier. The calculated DI and DBI of the clustering on the basis of day as a whole are 0.821 and 0.189, respectively, indicating a high degree of clustering on the basis of experimental day.

Following the removal of outliers, a distinction between the two days is still clearly present as depicted in Figure 6.13. While still significant, both the DI and DBI show a slight weakening: the DI decreases from 0.821 to 0.710 and the DBI increases from 0.189 to 0.241. Given the small changes in the validation measures, we can conclude that outlier removal alone was not enough to remove the statistical significance of temporal bias.

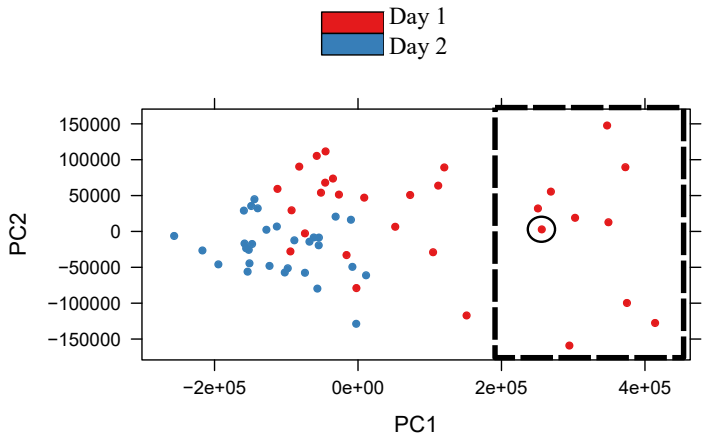

Figure 6.12: PCA Visualization of the Breast Cancer Dataset Before the Removal of Outliers, Grouped by Experimental Day. Before the removal of outliers, there exists a clear and distinct clustering on the basis of experimental day (Dunn index: 0.821, Davies–Bouldin index: 0.189). Arrays from the experiment on day 1 are represented by red points and the arrays from the experiment on day 2 are represented by blue points. The nine outliers and one near-outlier are highlighted by the dashed-line box on the right side, with the single near-outlier shown in a circle within that box.

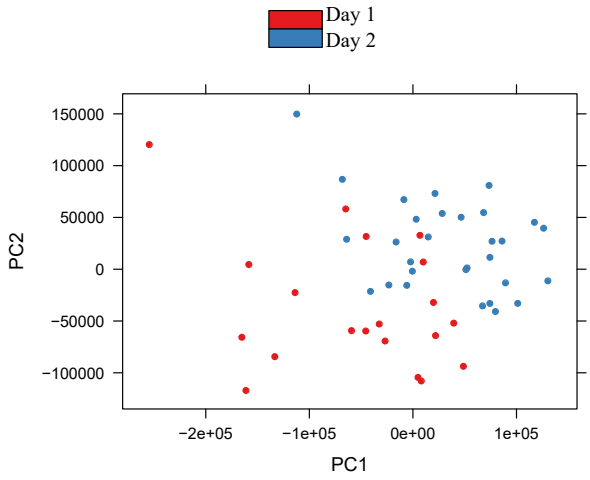

Figure 6.13: Breast Cancer Dataset Following the Removal of Outliers, Grouped by Experimental Day. After removal of outliers from the breast cancer dataset, there remains a very distinct relationship between groups of arrays and the day of experiment based on clustering indices. However, the DI and DBI have both become weaker, indicating that this distinction has reduced slightly, showing that temporal bias has been reduced, even if slightly (Dunn index: 0.710, Davies–Bouldin index: 0.241). Arrays from the experiment on day 1 are represented by red points and the arrays from the experiment on day 2 are represented by blue points.

Within day 1 taken alone, the results are much more promising. When outliers are included, the clustering is weak, which can be seen visually from the PCA as well as by the DBI/DI scores of .202 and 0.894, respectively (Figure 6.14A). When outliers are removed, the clustering is de-convoluted (Figure 6.14B). This is corroborated by the improved DI of 0.566 and an improved DBI of .210 (recall that a higher DI is indicative of better clustering while in contrast, a lower DBI is indicative of better clustering). Perhaps the most interesting aspect of these PCA graphs are the visual placement of the clusters. The samples containing the benign samples (blue) and the clusters containing the healthy samples (green) are now clustered together, while the cancer samples are much more spread.

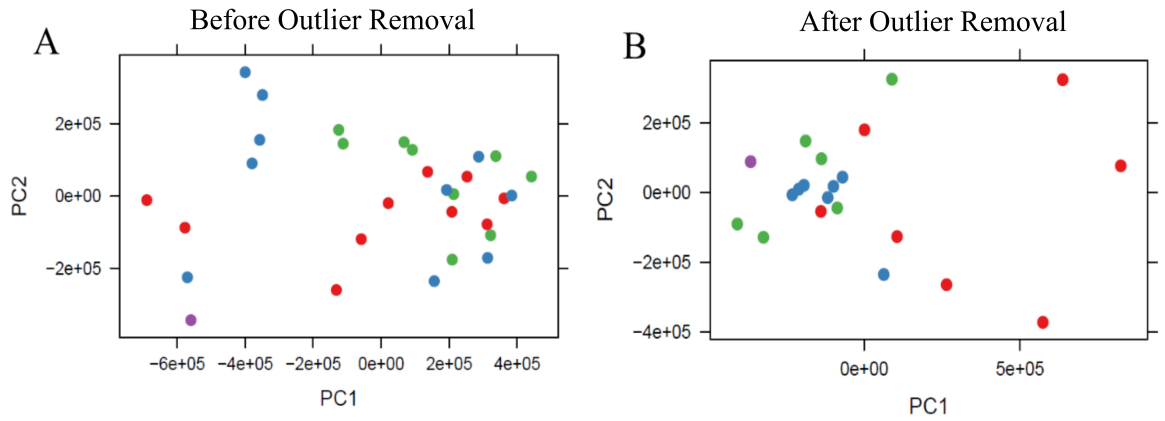

Figure 6.14: Breast Cancer Day 1 Samples Before (A) and After (B) Outlier Removal. Prior to outlier removal, it is difficult to distinguish a trend in the clustering of the arrays by sample. This is quantitatively expressed by the weak DI/DBI of 0.202/0.894. However, once the outlying 9 arrays are removed, a pattern emerges: the arrays from the healthy (green), papilloma (purple), and benign (blue) samples cluster tightly together, while the arrays from the cancer samples (red) form a very spread cluster. Quantitatively, the DI is improved to 0.566 and the DBI is improved to 0.210.

The last analysis that was conducted on the breast cancer dataset was analysis of the specific probes that were highly phosphorylated and their association with biological process and the similarities between those biological process across three categories: day 1 before outlier removal, day 1 after outlier removal, and day 2 (recall that day 2 had no outlying array). In this pursuit, PIIKA 2.5 was ran for the data from day 1 of the experiment before and after the removal of outliers and on the data from day 2 of the experiment. A qualitative comparison was performed between the lists of the 25 peptides with the highest fold increase in phosphorylation, as determined by PIIKA 2.5 to be consistently phosphorylated through both the X^2 test, for technical replicates, and the F test, for biological replicates (see [3]). These results are summarized in Table 6.6 where the lists of the 25 peptides with the greatest increase in phosphorylation are compared between the three datasets. From this table, it can be seen that both of the day 1 samples (before and after outlier removal) share two peptides in their top 25 list, day 1 before the removal of outliers and day 2 share only one peptide, and day 1 after outlier removal and day 2 share five peptides between them. When outliers are removed from day 1, the results are more similar to the results of day 2, which had no outliers. This implies that when no outliers are present, the patterns in phosphorylation level are more consistent, and suggests a benefit from the removal of outliers.

A more thorough effort could be made to identify which of these targets have the most significant associations with breast cancer through extensive literature review, but this is mostly past the scope of this thesis. Instead, a more holistic approach is taken here. The 25 peptides with the greatest difference in phosphorylation level between the healthy samples and a search was performed using those peptides for enrichment of specific biological processes using the online Gene Ontology Resource (GO). All GO searches were performed with the same settings, and p-values were corrected using the Bonferroni correction. Without the removal of outliers, the results of GO enrichment analysis are unsatisfactory. Only three terms were found

to be significant enriched: response to peptide hormone (p = 3.90E-02), response to organic substance (p = 3.61E-02), and cellular response to organic substances (p = 1.190E-02) (Table 6.7). Furthermore, these three terms had p values that were close to the significance threshold of p <0.05. In stark contrast, once the outliers were removed, the number of significantly enriched GO terms increased to 67 terms, including many terms highly relevant to phosphorylation, cellular signalling, and cancer (Table 6.8). This is much more similar to the results of GO enrichment analysis of the results from the day 2 arrays, in which 49 GO terms were found, including many of the same terms as day 1 following outlier removal (Table 6.9). There is a great degree in similarity between the two lists, with the most significant being the presence of a large number of phosphorylation-related terms with very high fold changes and very low p values. This speaks directly to PIIKA 2.5's ability to improve downstream analyses and interpretation of experimental results through improved quality control. Before the removal of outliers, the results provided by GO are practically a dead end, given they are incredibly general and only consist of three GO terms. In contrast, the wealth of relevant GO terms enriched through analysis after outlier removal is not only much more interpretable, but also provides many different directions for future research.

	Day 1, Outliers Included		Day 1, Outliers Removed		Day 2	
Rank	Target Location	P up	Target Location	P up	Target Location	P up
1	LAT	0.00E+00	CALD1	0.00E+00	APP	0.00E+00
2	PSTPIP1	0.00E+00	BTK	0.00E+00	CEBPA	0.00E+00
3	DGAT1	0.00E+00	KDR	1.00E-05	CCND1	0.00E+00
4	CD3G	0.00E+00	ACO2	2.00E-05	KIF15	0.00E+00
5	HSPA1A	0.00E+00	NFKB2	2.00E-05	F5	0.00E+00
6	CREB1	1.00E-05	GRK5	4.00E-05	CDK1	2.00E-05
7	DES	1.00E-05	CAMK2D	5.00E-05	MAP2K3	2.00E-05
8	IGFBP2	1.00E-05	PRKCE	6.00E-05	EZR	2.00E-05
9	DUSP1	1.00E-05	ITGA6	6.00E-05	GRK5	2.00E-05
10	IL12B	1.00E-05	NFE2L2	6.00E-05	SRSF1	2.00E-05
11	TGFBR1	1.00E-05	GRB10	9.00E-05	GIT2	3.00E-05
12	GSK3B	2.00E-05	DAPK1	1.00E-04	BLNK	3.00E-05
13	RAG2	5.00E-05	ABCC1	1.00E-04	ACSL4	3.00E-05
14	SRSF1	5.00E-05	CCND1	2.20E-04	STAT5B	4.00E-05
15	DPYSL3	6.00E-05	IQGAP1	2.20E-04	MAP2	5.00E-05
16	LPIN2	6.00E-05	CDK1	2.40E-04	ZAP70	5.00E-05
17	HSPH1	1.00E-04	EGFR	2.50E-04	SIRT2	1.10E-04
18	ACO2	1.30E-04	NCOA1	3.00E-04	PIKFYVE	1.30E-04
19	LPIN3	1.50E-04	TRIL	3.80E-04	MAP2K5	1.50E-04
20	UBQLN4	1.90E-04	TRAF6	4.00E-04	PRKCE	1.60E-04
21	AGPAT4	2.00E-04	NFATC2	5.30E-04	CAMK2A	1.80E-04
22	CDK12	2.10E-04	RPTOR	5.90E-04	RPS6KB1	1.80E-04
23	EIF2AK3	2.10E-04	FZD1	6.30E-04	L1CAM	2.00E-04
24	CROT	2.20E-04	PTPRCAP	7.90E-04	BTK	2.40E-04
25	AKT3	2.70E-04	CREB1	8.90E-04	STK10	2.80E-04

Table 6.6: Top 25 Peptides with the Highest Fold Increase in Phosphorylation Before and After Outlier Removal. Colour fields of cells indicate commonalities between columns. Blue, common between "Day 1, Outliers Included" and "Day 1, Outliers Removed"; Red, common between "Day 1, Outliers Included" and "Day 2"; green, common between "Day 1, Outliers Removed" and "Day 2". p values are calculated using the method provided by PIIKA[3]

GO Biological Process	Fold Change (+)	p value
cellular response to organic substance (GO:0071310)	4.88	1.19E-02
response to organic substance (GO:0010033)	3.96	3.61E-02
response to peptide hormone (GO:0043434)	13.61	3.90E-02

Table 6.7: GO Enrichment Analysis Results of the Day 1 Breast Cancer Data Before Outlier Removal. All fold change values are positive, representing the increase in fold change. Table is sorted by p value, smallest to largest, as calculated by Fisher's exact test with the Bonferroni correction for multiple hypothesis testing.

GO Biological Process	Fold Change (+)	p value
intracellular signal transduction (GO:0035556)	7.63	2.64E-06
signal transduction (GO:0007165)	3.54	3.32E-06
cellular response to organic substance (GO:0071310)	6.1	9.51E-06
signaling (GO:0023052)	3.31	1.27E-05
cell communication (GO:0007154)	3.24	1.92E-05
cellular response to chemical stimulus (GO:0070887)	5.04	2.90E-05
phosphorylation (GO:0016310)	9.87	3.40E-05
positive regulation of cellular metabolic process (GO:0031325)	4.48	3.59E-05
response to stimulus (GO:0050896)	2.41	3.81E-05
protein phosphorylation (GO:0006468)	11.45	5.40E-05
cellular response to stimulus (GO:0051716)	2.76	8.93E-05
positive regulation of metabolic process (GO:0009893)	3.84	9.53E-05
regulation of cell death (GO:0010941)	6.47	1.16E-04
cellular response to oxygen-containing compound (GO:1901701)	8.55	1.48E-04
regulation of response to stimulus (GO:0048583)	3.67	1.98E-04
response to chemical (GO:0042221)	3.65	2.20E-04
response to xenobiotic stimulus (GO:0009410)	15.58	2.52E-04
negative regulation of cellular process (GO:0048523)	3.31	2.84E-04
cellular response to chemical stress (GO:0062197)	21.12	2.87E-04
regulation of apoptotic process (GO:0042981)	6.73	3.60E-04
regulation of programmed cell death (GO:0043067)	6.6	4.50E-04
positive regulation of biological process (GO:0048518)	2.74	4.85E-04
response to organic substance (GO:0010033)	4.57	5.04E-04
response to oxygen-containing compound (GO:1901700)	6.37	6.59E-04
positive regulation of cellular process (GO:0048522)	2.9	6.96E-04
protein autophosphorylation (GO:0046777)	26.14	9.40E-04
response to inorganic substance (GO:0010035)	12.36	1.45E-03
response to metal ion (GO:0010038)	15.88	1.92E-03
negative regulation of metabolic process (GO:0009892)	4.13	1.94E-03
negative regulation of biological process (GO:0048519)	2.94	2.08E-03

Table 6.8: GO Enrichment Analysis Results of the Day 1 Breast Cancer Data After Outlier Removal. All fold change values are positive, representing an x-fold increase in the enrichment of terms in comparison to the expected value, where x is the fold change value. Values are sorted by p value, smallest to largest, as calculated by Fisher's exact test with the Bonferroni correction for multiple hypothesis testing. Only the 30 most significant results are shown. Total number of significant results: 67.

GO Biological Process	Fold Change (+)	p value
protein phosphorylation (GO:0006468)	14.89	4.19E-09
phosphorylation (GO:0016310)	11.66	8.74E-08
phosphate-containing compound metabolic process (GO:0006796)	6.22	3.83E-05
intracellular signal transduction (GO:0035556)	7.09	3.90E-05
phosphorus metabolic process (GO:0006793)	6.13	4.59E-05
positive regulation of growth (GO:0045927)	22.17	2.07E-04
cellular developmental process (GO:0048869)	3.95	2.55E-04
regulation of growth (GO:0040008)	11.82	2.92E-04
peptidyl-amino acid modification (GO:0018193)	9.34	3.68E-04
regulation of multicellular organismal process (GO:0051239)	4.49	6.34E-04
response to oxygen-containing compound (GO:1901700)	6.37	6.59E-04
developmental process (GO:0032502)	2.9	7.12E-04
cell cycle (GO:0007049)	7.25	7.96E-04
system development (GO:0048731)	3.65	8.72E-04
regulation of developmental growth (GO:0048638)	17.79	9.00E-04
protein autophosphorylation (GO:0046777)	26.14	9.40E-04
multicellular organismal process (GO:0032501)	2.63	1.11E-03
anatomical structure development (GO:0048856)	3.04	1.19E-03
cellular response to organic substance (GO:0071310)	5.28	1.29E-03
positive regulation of cell cycle (GO:0045787)	16.33	1.59E-03
positive regulation of developmental process (GO:0051094)	6.77	1.61E-03
cellular response to oxygen-containing compound (GO:1901701)	7.78	2.01E-03
cell differentiation (GO:0030154)	3.74	2.14E-03
cellular response to chemical stimulus (GO:0070887)	4.41	3.05E-03
regulation of cell cycle (GO:0051726)	7.34	3.41E-03
positive regulation of mitotic cell cycle (GO:0045931)	34.03	3.53E-03
protein modification process (GO:0036211)	4.34	3.72E-03
multicellular organism development (GO:0007275)	3.31	3.78E-03
response to organic substance (GO:0010033)	4.26	4.61E-03
mitotic cell cycle (GO:0000278)	10.51	4.90E-03

Table 6.9: GO Enrichment Analysis Results of the Day 2 Breast Cancer Data. All fold change values are positive, representing an x-fold increase in the enrichment of terms in comparison to the expected value, where x is the fold change value. Values are sorted by p value, smallest to largest, as calculated by Fisher's exact test with the Bonferroni correction for multiple hypothesis testing. Only the 30 most significant results are shown. Total number of significant results: 49.

In summation, the detection and removal of outliers from the breast cancer data yielded several distinct benefits. Firstly, the unwanted propensity for arrays to cluster by day was reduced, even if slightly. Ultimately, the DI decreased from 0.821 to 0.710 and the DBI increased from 0.189 to 0.241, a small weakening of both values. Through this, it was also discovered that a large source of this error was the fact that all of the outliers and near-outliers detected by PIIKA 2.5 were detected on day 1, though this could not account for the entire error. Secondly, removal of outliers increased the clustering on the basis of sample phenotype (cancer, benign, or healthy) significantly. This yielded a pattern that reflected the nature of breast cancer and cancers in general. The healthy and benign samples clustered together very tightly on the basis of their phosphorylation intensities, while the cancer samples were very spread out, indicating a diverse pattern of phosphorylation phenotypes. Thirdly, the removal of outliers increased the similarity between the targets with the most upregulated phosphorylation. Before removal of outliers, only one target was similar between the two days; after removal, this increased to five. Lastly, the results of Gene Ontology analysis were far more extensive, as well as more relevant to the biological issues in question, directly indicating that the removal of outliers leads to more interpretable results and an increase in the number of potential avenues of research, exemplifying the value of PIIKA 2.5.

6.5 Application of All Measures on a Single Dataset

Though the material described in the previous sections focuses on the validation of the new quality control measures in PIIKA 2.5, both datasets have been manually realigned and did not require spot size corrections. This section remedies this by assessing all of PIIKA 2.5's quality control features in combination through their application on a dataset with none of PIIKA 2.5's features applied (hereafter referred to as the "before" dataset) and then again following the application of all three (hereafter referred to as the "after" dataset). As aim of this is to highlight the differences between an experiment with no quality control to an experiment that utilises all of the features of PIIKA 2.5 to their fullest extent, all corrections suggested by PIIKA 2.5's results (included those suggested by amber lights) are applied.

The dataset that was selected for this process was again from the LPS–cortisol experiment, but using the fluorescent-labelled ATP subsection prior to any manual correction. This was selected primarily because the data are from an experiment which is based on the well-characterized interaction between cortisol, LPS, and the innate cellular immune system. Additionally, the dataset has the key characteristic necessary for this analysis: all three quality control improvements are applicable to it (other datasets had pre-existing quality such that at least one of PIIKA 2.5's quality control measures was not needed). A thorough description of the dataset's creation and characteristics can be found in Section 6.1.1.

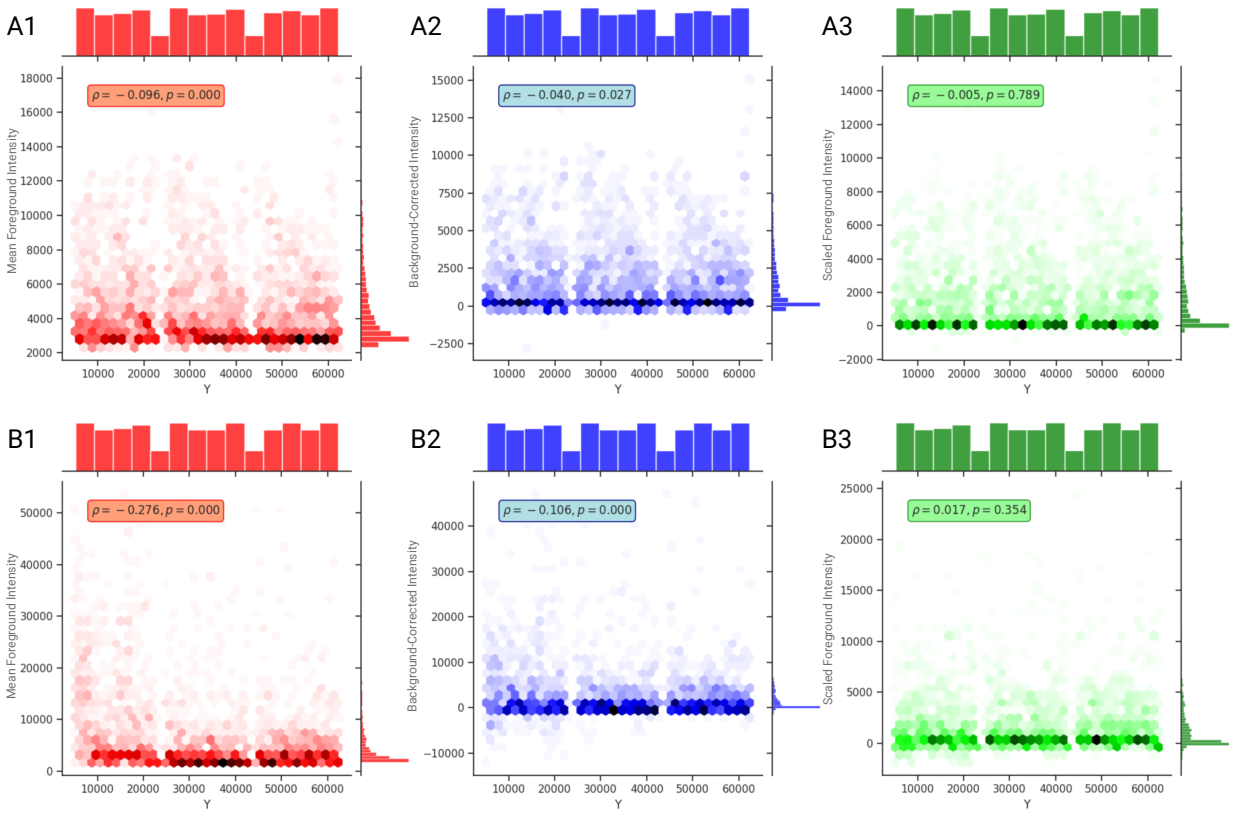

Figure 6.15: Two Examples of Background Correction Methods Applied to the Labelled ATP Dataset

6.5.1 Application of Background Scaling

Background correction was applied to both the "before" and "after" data subsets. In the case of the "before" dataset, background subtraction was performed, as is typical. In the case of the "after" dataset, background scaling was performed. Two examples of the effectiveness of background scaling in this experiment are demonstrated in Figure 6.15. Background scaling provided a superior reduction in location bias than background subtraction in 25 out of 27 arrays (92.6%), based on the Spearman ρ value calculated for the correlation between y-axis position on the array and fluorescent intensity. A full comparison of the use of background scaling, background subtraction, and no background correction on this dataset was shown in a previous section Table 6.2.

Figure 6.16: Summary of Realignment of Data. Arrays (n = 27) were aligned using GenePix's automatic alignment feature. Of these arrays, all but five exhibited a mean–median intensity slope of less than 0.75, with one of those arrays exhibiting a slope of 0.783 (array 20). Manual correction of misaligned blocks was performed on arrays 4, 10, 20 and 21. Realignment was performed, with spot size adjustment. Following this, all arrays exhibited a mean–median slope greater than 0.75, with three arrays having slopes between 0.75 and 0.80. For these three arrays, a secondary realignment was performed.

6.5.2 Correction of Improper Spot Size

The typical workflow at the experimental stage immediately following image capture utilises the automatic alignment algorithm provided by the scanner vendor to determine the sizes and position of the spots. However, this sometimes involves manual adjustment in cases where obvious errors in array alignment are present. In the creation of the "before" dataset, this typical workflow was followed, and across the 27 arrays in the dataset, only 4 required manual correction due to obvious alignment errors. All four cases occurred where an entire block (a third of an array) was misaligned greatly, causing most probes of that block to be completely missed by the grid.

Of the arrays from the output from GenePix, only four received a green light from PIIKA 2.5's spot size evaluation feature. This assessment was performed after background correction through PIIKA 2.5's background scaling method. In total, 22 arrays received a red light from PIIKA 2.5's spot size evaluation feature. An additional array received an amber light (slope between 0.75 and 0.80, specifically in this case: 0.783). Based on these results, manual realignment was performed for all 22 red-flagged arrays and the single amber-flagged array. Following manual realignment, all but 3 arrays (24 arrays) were improved to have a mean–median slope greater than 0.80; those three arrays all received an amber light. Realignment was then performed a second time on the three amber light arrays. Given that these arrays were already manually realigned, some researchers may wish to skip this and proceed with an amber-flagged array. This is a perfectly acceptable choice in such a case. However, to fully demonstrate the quality control features of PIIKA 2.5, all amber light results are included in this demonstration. A summary of the results of this process is presented in Figure 6.16.

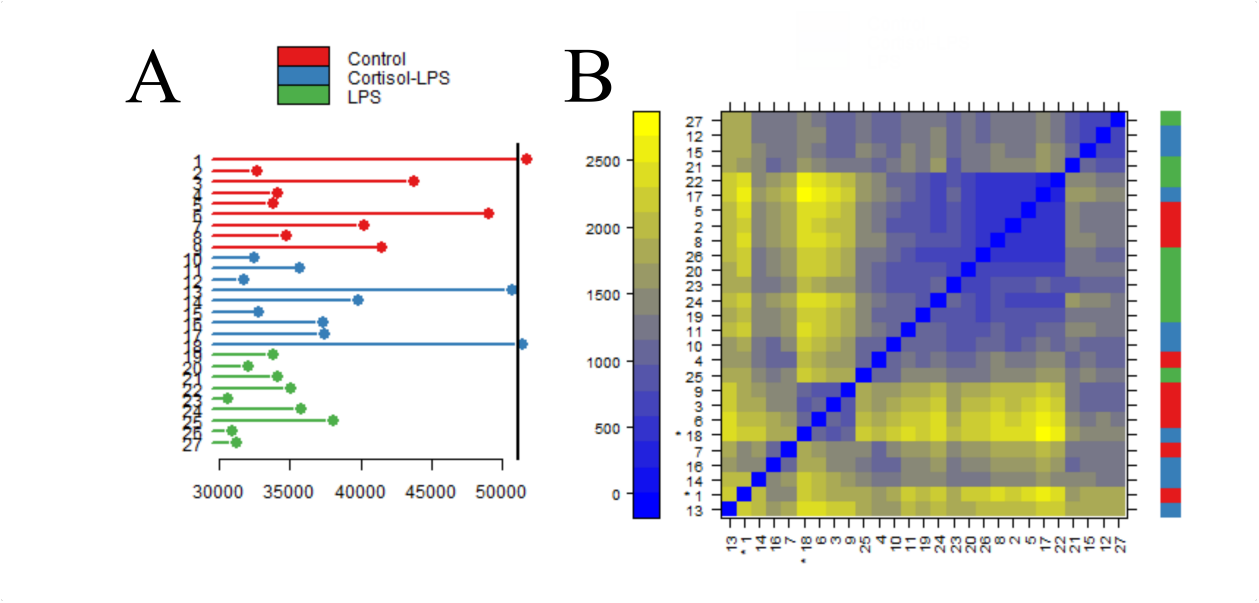

Figure 6.17: Outlier Detection on the Labelled ATP Subset of the LPS Stimulation Dataset Outlier detection determined the presence of two outliers: arrays 1 and 18. Arrays 5 and 13 were identified to be near-outliers. All outliers and near-outliers were removed prior to downstream analysis.

6.5.3 Outlier Detection and Removal

Outlier array detection and analysis was performed as described in previous sections (Section 6.3 and Section 6.4). It was determined that two outliers were present (p <0.05), array 1, a control group array from experimental day 1, and array 18, a Cortisol + LPS array from experimental day 3 (Figure 6.17). In addition, two more arrays (arrays 5 and 13) were near-outliers (p <0.1). For this demonstration, the near-outliers were also removed from the dataset.

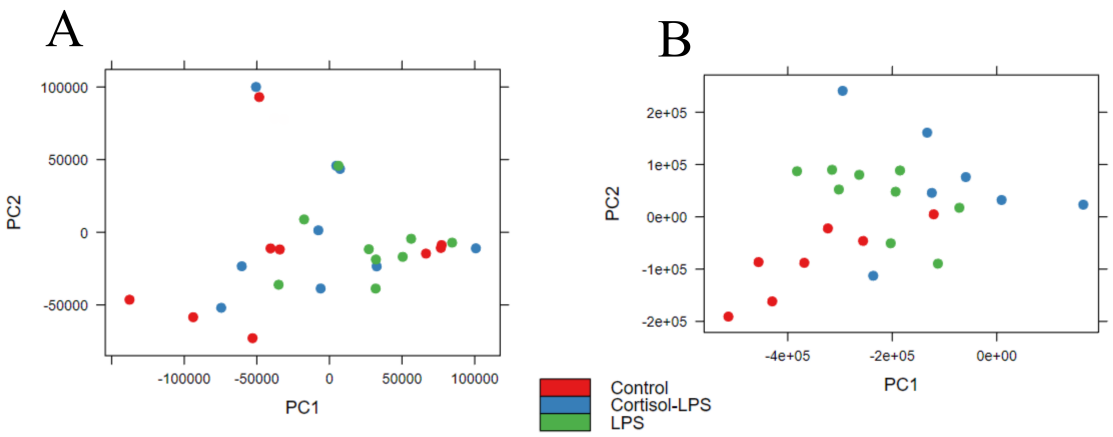

Figure 6.18: Results After All Measures are Applied to a Single Dataset. A. PCA and calculated DI/DBI values before any of PIIKA 2.5's new quality control improvements have been applied. Arrays were aligned using GenePix automatic alignment with manual adjustment only when features were egregiously out of place, such as when entire blocks were shifted. No resizing of spots was performed in any case. Note that background subtraction was still performed as is typical. n = 27, with nine arrays each from three treatment groups (control, red; cortisol and LPS, blue; LPS only, green). The results show poor distinction between the three treatment groups, which is reflected by the DI and DBI values of 0.161 and 0.832, respectively. **B.** PCA and calculated DI/DBI values after all of the major quality control improvements are applied. Realignment of the array was performed Substituting background subtraction, PIIKA 2.5's new background scaling method is used instead. n = 23, with identical characteristics to A. with the exception of the removal of four arrays, two each from the control and cortisol-LPS treatment groups. The cluster validation metrics are highly improved, to 0.690 and 0.149 for DI and DBI, respectively. This provides strong evidence for the application of PIIKA 2.5's quality control measures.

6.5.4 Results

When compared to the results prior to any quality control measures there is marked improvement in the quality of the clustering after all three quality control measures are applied (Figure 6.18). In this analysis, the Dunn index (DI) is improved from 0.161 to 0.690 and the Davies–Bouldin index (DBI) is improved from 0.832 to 0.149. These are substantial improvements to both cluster validation metrics. Background subtraction was still performed on the "before" results, as this option is typically performed in most experiments and available in PIIKA 2. Besides the quantitative measures, it is also visually apparent that the distinction between treatments has substantively been improved by the application of PIIKA 2.5's quality control measures. Overall, this result provides strong evidence of the usefulness of PIIKA 2.5's quality control measures.

6.6 Discussion

Quality control analysis for microarray programs in several popular R packages has led to increased reproducibility [83] [84] [85], one of the most important goals in modern high-throughput data analysis. Given the challenges presented by the peptide array technology, it was desirable to have similar quality control filters

integrated into PIIKA 2 that specifically targeted major identifiable problems.

One of the most apparent issues in kinome quality control is the lack of a dedicated kinome microarray imaging software. Using automatic grid alignment for the microarrays creates an error caused by mislabelling pixels. One solution is manual realignment of the grid in the imaging software, but this is time-consuming and prone to error. Unless specifically told the proper way to estimate spot size 4.1, manual size adjustment can just as easily create the same mislabelling error.

By comparing the mean and the median of each probe after background correction we can get a measurement of the symmetry of the data (Figure 4.2). In a situation with proper pixel labelling, it is expected that the distributions of the means and medians of probes are roughly equivalent (Figure 4.2B), but when the foreground intensity is highly skewed (resulting in the median and mean distributions being dissimilar (Figure 4.2A2)) it indicates that background pixels were captured within the area labelled as foreground (Figure 4.1C). This is often easily recognizable by the characteristic fin shape and low mean–median slope (Figure 4.3), but the effects may be more obscure but still detrimental. The probes that lie in the fin area of the graph (typically between mean 5000–15000 and median <5000) have been significantly shifted down in both background-corrected foreground mean and background-corrected foreground median. These probes are relevant as, if the data were better conditioned, it could be that these probes would have much higher overall intensities and be placed among the probes with the highest levels of phosphorylation—the "top hits" in a differential phosphorylation analysis.

Background scaling is an important first step in the correction of location biases. We have discovered that in kinome arrays there exists a systematic bias that causes the intensities to increase as the y-axis value increases (Figure 6.5). This bias exists uniformly across all 27 arrays analysed in the LPS stimulation experiment and is largely ubiquitous across all microarray technologies [36, 86]. This is a result of the reagent used to selectively stain for phosphorylated peptides being increasingly intense towards one end of the array. Location bias is not adequately corrected with background subtraction alone (Figure 6.54B). In two-color arrays, many different background correction techniques are used, and have been rigorously analysed [87]. Methods similar to the one presented here have been effective in reducing false positives associated with location-associated bias without compromising differential expressed genes in two-color arrays [86]. However, in kinome microarrays, there are fewer options for background correction algorithms due to the relative scarcity of kinome microarray analysis software. We have shown that the location-associated variation can be further reduced by scaling the foreground and background with a local estimate of relative background location bias prior to the background subtraction (Figure 6.5C). Our method is presented as an alternative to background subtraction, adding to existing corrections which also outperform background subtraction in reduction of location-associated biases [88]. Calibration probes and other quality control standards which have been previously unsuccessful due the level of noise between technical replicates can be reinvestigated with the application of background scaling. Future efforts can be made to incorporate calibration probes to eliminate a wider variety of biases that can obfuscate the experimental results now that the location-associated

variation can be more adequately corrected for.

Pairwise comparison of all the probes on each array is an effective tool for measurement of how similar arrays are. The distance between all corresponding probes between two arrays is measured and the mean of all distances from one array to all other arrays determined. Using an ANOVA, the outliers are determined as arrays with a sum of pairwise distances to all other arrays significantly different from the sum for other arrays. In-house trials have indicated that arrays that are outliers are variant by a large margin (well beyond the outlier threshold) while the non-outlying arrays are very consistent. Being an outlier therefore indicates presence of a factor that affects all of the probes in the array systematically, as opposed to just a particular subset of probes (the highest intensity probes, for example). There may be situations where the outlier arrays represent a biological effect, so it is necessary for researchers to exercise judgement on how to best proceed. In many cases, however, differentially phosphorylated peptides between treatment groups represent only a small portion of the total number of peptides in the array, so it is more likely that identification as an outlier is caused by sweeping systematic biases.

The quality control results are supplied to the user in a folder with the other data that PIIKA provides, containing the analysis of the two metrics described here. The program supplies figures similar to Figure 4.3 and 6.6 that show the mean–median slopes for each array analysed and the determined outlier arrays. A three-tiered stoplight system has been implemented to further guide researchers in making decisions based on the data provided. This system is ultimately subjective but uses objective data to categorize data into three categories: green (having no observable problems), red (having a large detectable issue), and amber (potentially having some issues) based on the criteria described in the Materials and Methods. These colours are integrated into the provided figures and more detailed explanations are offered in an accompanying text file. Ideally, it would be possible to categorize the data into just two categories, but the amount of variation in the data creates sufficient ambiguity to make a binary approach impossible. Green quality assessments mean that there was no problem found with the data at all, giving the user more confidence that their data are reliable. Red assessments are accompanied by a message that strongly suggests some correction to the data or complete elimination is required. There is also a third category, amber, for datasets that are not distinctly "bad" nor distinctly "good". Researchers could still use these datasets effectively with caution, but should downstream analysis become problematic, a potential source of error has been identified.

Despite this being a major improvement to the existing kinome microarray analysis pipeline, there are still challenges that have yet to be addressed. The lack of image analysis software specifically tailored for kinome microarrays is a major limitation and addressing this would mitigate many of the problems that are corrected by this update. A dedicated software suite would also alleviate much of the time-consuming labour involved in manual alignment of the array image with the data grid. While background scaling does provide correction for spatial biases, the update to PIIKA does not correct for other systematic biases that we have determined to be evident in kinome microarrays—it only identifies them. It is feasible that eventually problems such as misalignment of the imaging grid, batch bias, and overall array quality will not only be identifiable, but

also correctable. Regardless of this, the identification of the problems and their sources is the essential first step of correction. The update to PIIKA provides objective information for researchers to make subjective decisions about their data, which will lead to increased accuracy and technical reproducibility.

7 Conclusions

7.1 Conclusion

As kinome analysis has not proliferated as widely as genomic and transcriptomic analysis, there are not as many computational tools to enhance the experimental results from kinome arrays so as to improve data quality and ease data interpretation. In this work, a new addition to PIIKA, PIIKA 2.5, was developed and applied to enhance the results of kinome analysis experiments.

Through literature review and independent research, key quality control areas were identified, namely background correction and outlier detection, which are relevant issues in genomics and transcriptomics, and spot size assessment, which has not previously been identified in other similar array technologies. For the prior two areas, it was possible to leverage methods developed for these other technologies for the benefit of kinome analysis. For the latter, exploration was required to develop the means by which to identify the erroneous pattern created by excessively large spot size. The overall objectives of this thesis were to develop these methods, incorporate them into the existing kinome analysis pipeline, and validate them through several case studies.

A novel background correction was developed, utilising the median intensity of the array to scale the background signal before subtraction. This background correction was originally proposed by Conor Lazarou, and implemented here. Through validation on several different datasets, this new method, background scaling, was identified to be superior to the most commonly used background correction method, background subtraction. This feature was successfully developed and incorporated into PIIKA as an additional option for users to choose from. In all datasets analysed, background scaling had a larger reduction in the correlation between physical position on the array and probe intensity value and removed this bias entirely in more arrays than background subtraction.

A method for detection of error caused by improper spot sizing of automated algorithms was developed, by analysing the features of data corrupted in this fashion in comparison to manually corrected arrays. It was found that the skewness of the data (the difference between mean and median pixel intensity for each probe) was a reliable metric on which to determine the arrays most significantly affected by this phenomenon. Application of manual resizing, although an arduous task, results in the elimination of the error caused by the commonly used automated algorithms built into the imaging software.

An outlier detection method was chosen and adapted from literature for nucleic acid microarrays out of several commonly used methods. The chosen method utilises the sum of pairwise distance between probes

from an array and all other arrays, with arrays that have a significantly higher sum of distances being flagged as outliers. This feature was implemented to provide users with three different classifications of arrays—outlier, near-outlier, and normal, presented as red, amber, and green flags, respectively—given alongside the normal results of PIIKA. This allows researchers to make informed decisions in future analysis. Through three separate investigations utilising two cluster validation measures, PIIKA 2.5's new outlier analysis was shown to increase the tendency of arrays to cluster by treatment group and reduces the tendency of arrays to cluster on the basis of unintended experimental factors that emerge in a date-dependent manner. Through these investigations, it was made clear that the removal of outliers facilitated easier interpretation of results and made the results of experiments more consistent with each other.

PIIKA 2.5 presents a major update to the kinome analysis workflow, bringing the bioinformatics technology closer to the level of the much more prolific and advanced methods for nucleic acid microarray analysis and improving the quality and consistency of experimental results emerging from peptide arrays for kinome analysis.

7.2 Limitations and Future Investigations

The project as currently presented has several limitations which can be corrected for with work that is outside the scope of this master's thesis. The first of these is an issue present with the method for determination of improper spot size, which causes an error that manifests in several different unusual statistical features. As previously discussed in Section 4.3, when the spot size is too large, as performed by either the automated alignment feature of the scanner vendor software or by manual alignment, an erroneous effect can be observed in the skewness (the difference between the median and mean) of the data caused by mislabelling of background pixels as foreground prior to the calculation of the summary statistics of the probes. This situation affects probes in an erratic fashion, significantly lowering both the median and mean values of the probe, but the median is affected much more. Certain high-intensity probes that this error affects become a part of a phenomenon referred elsewhere in this thesis as a crest or a fin (Section 4.3). This phenomenon is variant in presentation and was not a suitable indicator of spot size error. However, because the summary statistics of high intensity probes are altered more than those with lower intensities, and the same situation affects lower intensity probes, it creates an anomaly in a plot of foreground median versus foreground mean. This anomaly was detectable by determining the slope of the two values, as determined by linear regression. A slope significantly lower than one was shown to be indicative of excessively large spot size. This works well for the time being but has two limitations that could be solved in the future.

The first limitation with the slope measurement method is that the thresholds for classification using the three-tier "stoplight" system were determined qualitatively through repeated examination of multiple datasets. These thresholds were corroborated with data from another lab performing the same type of kinome microanalysis; even so, these thresholds have not been proven to be valid in all cases. Future investigation

on the appropriateness of these thresholds would involve an extensive analysis on the mean–median slope's characteristics across many datasets, including its mean, standard deviation, and overall distribution. After those summary statistics on the means are gathered, it would be possible to conduct an appropriate statistical test to find data-derived thresholds based on determined p values. A major caveat of this is that it will require a selection of arrays that have been very carefully manually aligned or the manual realignment of many arrays, which would be intensely time consuming.

The second limitation of the slope measurement method which could be solved in future investigations is motivated by an issue discovered late in this project: when a kinome microarray was intentionally aligned poorly, the slope method no longer functioned properly. Extremely poor alignment increases the density of the fin-shaped anomaly significantly, resulting in an increase in the value of the mean–median slope, counteracting the overall slope decrease that is used to identify misalignment. In the severest cases, the mean–median slope can even become greatly above 1. This issue cannot be automatically detected with the automated measure proposed in this thesis and requires manual inspection to identify. In these cases, the fin-shaped area in the median–mean plots is so exaggerated that it would be easy to identify by an aware researcher. While this seems like a significant issue, it has only been discovered in arrays that have been intentionally manipulated to be as poorly aligned as is reasonably possible, much worse than the automatic detection algorithm.

Ultimately, both of the previously discussed possibilities naturally lead into a possibility for correction of the problems entirely. Whether this is even possible is difficult to determine. For correction of the spot size error based on output probe intensities, an approach would involve analysis of a large number of datasets with this error to develop a mathematical model of how this error precisely affects the output probe intensities. Although this thesis did examine many datasets, they were not examined with this intention in mind, and it is likely that the ultimate number of datasets would have to be so large as to be infeasible. Additionally, a strong possibility exists that the spot size effect is so variable that it cannot be modelled in such a way. Another approach to correct the error caused by excess spot size is much broader, involving the prevention of the error from arising in the first place through the use of a grid detection algorithm specifically optimized for use in kinome experiments. If this problem cannot be corrected for, an alternative to the current method could be the use of quality scores for individual probes. For example, next-generation sequencing technologies often use the Phred score, which represents the likelihood of finding an incorrectly sequenced base [89]. This contributes to a broader set of objective criteria that can be used to eliminate individual reads in sequencing experiments [90]. A similar per-spot measure has potential applications in kinome microarray technology, using a combination of the physical characteristics of the spots to create an aggregated score that may be able to eliminate arrays in a more objective manner or eliminate individual probes from the experiment instead of entire arrays, which would save a great deal of manual effort.

Furthermore, there could also be other combination of scanner/scanner software that need to be assessed to determine if the same error is present in their manual alignment algorithms. Because this thesis only

utilised data from the GenePix scanner produced by the company Molecular Devices, the results presented here may limited to this analysis platform. Future investigation is required in this regard, as switching experimental set ups to a platform that does not create this error in its automatic alignment would save a large amount of manual labour.

Correction for the issue of outliers is also a possibility in the future, but is likely to be even harder to undertake than spot size error correction, especially as the potential sources that influence whether an array is an outlier or not have not been identified. It is difficult to speculate on the path which by outlying arrays could be corrected to make them coexist with the other arrays without knowing the source of error that led to classification. Several normalisation measures and, perhaps, additional background correction methods could be explored in pursuit of rectifying some outlying arrays. Furthermore, the background scaling method presented in this thesis could be compared against other forms of background correction.

Further work could be conducted through application of PIIKA 2.5's features on many more datasets, including those from other labs that may use different protocols. The measures presented in this thesis could be validated through more analyses, especially those that utilise all three features to their fullest extent, such as in Section 6.5. Lastly, PIIKA can be advanced through other means, not necessarily in the realm of quality control. Additions to PIIKA could incorporate powerful new advances in machine learning technology into its analysis.

References

[1] Gerard Manning, David B Whyte, Ricardo Martinez, Tony Hunter, and Sucha Sudarsanam. The protein kinase complement of the human genome. *Science*, 298(5600):1912–1934, 2002.

[2] Antonio Facciuolo, Connor Denomy, Sean Lipsit, Anthony Kusalik, and Scott Napper. From beef to bees: High-throughput kinome analysis to understand host responses of livestock species to infectious diseases and industry-associated stress. *Frontiers in Immunology*, 11, May 2020.

[3] Yue Li, Ryan J. Arsenault, Brett Trost, Jillian Slind, Philip J. Griebel, Scott Napper, and Anthony Kusalik. A systematic approach for analysis of peptide array kinome data. *Science Signaling*, 5(220):pl2, April 2012.

[4] Brett Trost, Jason Kindrachuk, Pekka Määttänen, Scott Napper, and Anthony Kusalik. PIIKA 2: An expanded, web-based platform for analysis of kinome microarray data. *PLoS ONE*, 8(11):e80837, November 2013.

[5] Connor Denomy, Conor Lazarou, Scott Napper, and Anthony Kusalik. Enhanced quality control for kinome microarrays, July 2020.

[6] Connor Denomy, Conor Lazarou, Daniel Hogan, Antonio Facciuolo, Erin Scruten, Anthony Kusalik, and Scott Napper. PIIKA 2.5: Enhanced quality control of peptide microarrays for kinome analysis. *Plos one*, 16(9):e0257232, 2021. Publisher: Public Library of Science San Francisco, CA USA.

[7] Conor Lazarou. Kinome array analysis location bias and target calibration, August 2016.

[8] Robert Roskoski. Properties of FDA-approved small molecule protein kinase inhibitors: A 2021 update. *Pharmacological Research*, 165, March 2021.

[9] Philip Cohen. The origins of protein phosphorylation. *Nature cell biology*, 4(5):E127–E130, 2002. Publisher: Nature Publishing Group.

[10] Akanksha Baharani, Brett Trost, Anthony Kusalik, and Scott Napper. Technological advances for interrogating the human kinome. *Biochemical Society Transactions*, 45(1):65–77, February 2017.

[11] Brett Trost, Ryan Arsenault, Philip Griebel, Scott Napper, and Anthony Kusalik. DAPPLE: a pipeline for the homology-based prediction of phosphorylation sites. *Bioinformatics*, 29(13):1693–1695, July 2013.

[12] Brett Trost, Farhad Maleki, Anthony Kusalik, and Scott Napper. DAPPLE 2: a Tool for the Homology-Based Prediction of Post-Translational Modification Sites. *Journal of Proteome Research*, 15(8):2760–2767, 2016. _eprint: https://doi.org/10.1021/acs.jproteome.6b00304.

[13] Philip Cohen. Protein kinases — the major drug targets of the twenty-first century? *Nature Reviews Drug Discovery*, 1(4):309–315, April 2002.

[14] Mark Schena, Dari Shalon, Ronald W Davis, and Patrick O Brown. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, 270(5235):467–470, 1995. Publisher: American Association for the Advancement of Science.

[15] Deval A Lashkari, Joseph L DeRisi, John H McCusker, Allen F Namath, Cristl Gentile, Seung Y Hwang, Patrick O Brown, and Ronald W Davis. Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proceedings of the National Academy of Sciences*, 94(24):13057–13062, 1997. Publisher: National Acad Sciences.

[16] Edwin M Southern. DNA microarrays. *DNA arrays*, pages 1–15, 2001. Publisher: Springer.

[17] Roger Bumgarner. Overview of DNA microarrays: types, applications, and their future. *Current protocols in molecular biology*, 101(1):22–1, 2013. Publisher: Wiley Online Library.

[18] Sungjin Park, Jeffrey C Gildersleeve, Ola Blixt, and Injae Shin. Carbohydrate microarrays. *Chemical Society Reviews*, 42(10):4310–4326, 2013.

[19] Li Feng. Probing lipid–protein interactions using lipid microarrays. *Prostaglandins & other lipid mediators*, 77(1-4):158–167, 2005.

[20] Ziqing Chen, Tea Dodig-Crnković, Jochen M Schwenk, and Sheng-ce Tao. Current applications of antibody microarrays. *Clinical proteomics*, 15(1):1–15, 2018.

[21] Aakash Chawade, Erik Alexandersson, and Fredrik Levander. Normalyzer: a tool for rapid evaluation of normalization methods for omics data sets. *Journal of proteome research*, 13(6):3114–3120, 2014. Publisher: ACS Publications.

[22] Wentian Li, Yaning Yang, SM Lin, and KF Johnson. Methods of microarray data analysis, 2002.

[23] K Shakya, HJ Ruskin, G Kerr, M Crane, and J Becker. Comparison of microarray preprocessing methods. In *Advances in Computational Biology*, pages 139–147. Springer, 2010.

[24] Sorin Drăghici. *Data analysis tools for DNA microarrays*. CRC Press, 2019.

[25] Attri Ghosal, Arunima Nandy, Amit Kumar Das, Saptarsi Goswami, and Mrityunjoy Panday. A short review on different clustering techniques and their applications. *Emerging technology in modelling and graphics*, pages 69–83, 2020.

[26] Yue Li, Ryan J. Arsenault, Brett Trost, Jillian Slind, Philip J. Griebel, Scott Napper, and Anthony Kusalik. A systematic approach for analysis of peptide array kinome data. *Science Signaling*, 5(220):pl2, April 2012.

[27] Antonio Facciuolo, Connor Denomy, Sean Lipsit, Anthony Kusalik, and Scott Napper. From beef to bees: High-throughput kinome analysis to understand host responses of livestock species to infectious diseases and industry-associated stress. *Frontiers in Immunology*, 11(May):789, 2020. Publisher: Frontiers.

[28] Famatta Perry, Casey Johnson, Bridget Aylward, and Ryan J. Arsenault. The Differential Phosphorylation-Dependent Signaling and Glucose Immunometabolic Responses Induced during Infection by Salmonella Enteritidis and Salmonella Heidelberg in Chicken Macrophage-like cells. *Microorganisms*, 8(7):1041, July 2020.

[29] Jason Kindrachuk, Victoria Wahl-Jensen, David Safronetz, Brett Trost, Thomas Hoenen, Ryan Arsenault, Friederike Feldmann, Dawn Traynor, Elena Postnikova, Anthony Kusalik, et al. Ebola virus modulates transforming growth factor β signaling and cellular markers of mesenchyme-like transition in hepatocytes. *Journal of virology*, 88(17):9877–9892, 2014.

[30] Gilles Rabatel, Federico Marini, Beata Walczak, and Jean-Michel Roger. VSN: Variable sorting for normalization. *Journal of Chemometrics*, 34(2):e3164, 2020.

[31] Jetse Scholma, Gwenny M. Fuhler, Jos Joore, Marc Hulsman, Stefano Schivo, Alan F. List, Marcel J. T. Reinders, Maikel P. Peppelenbosch, and Janine N. Post. Improved intra-array and interarray normalization of peptide microarray phosphorylation for phosphorylome and kinome profiling by rational selection of relevant spots. *Scientific Reports*, 6(1):26695, July 2016.

[32] Taesung Park, Sung-Gon Yi, Sung-Hyun Kang, SeungYeoun Lee, Yong-Sung Lee, and Richard Simon. Evaluation of normalization methods for microarray data. *BMC bioinformatics*, 4(1):1–13, 2003.

[33] Farnoosh Abbas-Aghababazadeh, Qian Li, and Brooke L Fridley. Comparison of normalization approaches for gene expression studies completed with high-throughput sequencing. *PloS one*, 13(10):e0206312, 2018.

[34] Jacob E Wulff, Matthew W Mitchell, et al. A comparison of various normalization methods for lc/ms metabolomics data. *Advances in Bioscience and Biotechnology*, 9(08):339, 2018.

[35] Daniel S. Yuan and Rafael A. Irizarry. High-resolution spatial normalization for microarrays containing embedded technical replicates. *Bioinformatics*, 22(24):3054–3060, October 2006. _eprint: https://academic.oup.com/bioinformatics/article-pdf/22/24/3054/16852177/btl542.pdf.

[36] KK Dobbin, Ernest S Kawasaki, DW Petersen, and RM Simon. Characterizing dye bias in microarray experiments. *Bioinformatics*, 21(10):2430–2437, 2005.

[37] Tristan Mary-Huard, Jean-Jacques Daudin, Stéphane Robin, Frédérique Bitton, Eric Cabannes, and Pierre Hilson. Spotting effect in microarray experiments. *BMC Bioinformatics*, 5(1):63, May 2004.

[38] Doris Steger, David Berry, Susanne Haider, Matthias Horn, Michael Wagner, Roman Stocker, and Alexander Loy. Systematic spatial bias in DNA microarray hybridization is vaused by probe spot position-dependent variability in lateral diffusion. *PLOS ONE*, 6(8):1–11, August 2011. Publisher: Public Library of Science.

[39] Gábor Balázsi, Krin A. Kay, Albert-László Barabási, and Zoltán N. Oltvai. Spurious spatial periodicity of co-expression in microarray data due to printing design. *Nucleic Acids Research*, 31(15):4425–4433, August 2003. _eprint: https://academic.oup.com/nar/article-pdf/31/15/4425/3865295/gkg485.pdf.

[40] Philippe Serhal and Sébastien Lemieux. Correction of spatial bias in oligonucleotide array data. *Advances in Bioinformatics*, 2013, 2013.

[41] Jin Hwan Do and Dong-Kug Choi. Normalization of microarray data: single-labeled and dual-labeled arrays. *Molecules & Cells (Springer Science & Business Media BV)*, 22(3), 2006.

[42] Gordon K. Smyth and Terry Speed. Normalization of cDNA microarray data. *Methods*, 31(4):265–273, 2003.

[43] Yee Hwa Yang, Sandrine Dudoit, Percy Luu, David M Lin, Vivian Peng, John Ngai, and Terence P Speed. Normalization for cdna microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic acids research*, 30(4):e15–e15, 2002.

[44] Frédéric Normandeau, Andy Ng, Maiwenn Beaugrand, and David Juncker. Spatial bias in antibody microarrays may be an underappreciated source of variability. *ACS sensors*, 6(5):1796–1806, 2021.

[45] Bogdan Mazoure, Robert Nadon, and Vladimir Makarenkov. Identification and correction of spatial bias are essential for obtaining quality data in high-throughput screening technologies. *Scientific reports*, 7(1):1–10, 2017.

[46] Margaret A. Taub, Hector Corrada Bravo, and Rafael A. Irizarry. Overcoming bias and systematic errors in next generation sequencing data. *Genome Medicine*, 2(12):87, December 2010.

[47] Jeffrey T. Leek, Robert B. Scharpf, Héctor Corrada Bravo, David Simcha, Benjamin Langmead, W. Evan Johnson, Donald Geman, Keith Baggerly, and Rafael A. Irizarry. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature Reviews Genetics*, 11(10):733–739, October 2010.

[48] Peter V Hornbeck, Indy Chabra, Jon M Kornhauser, Elzbieta Skrzypek, and Bin Zhang. PhosphoSite: A bioinformatics resource dedicated to physiological protein phosphorylation. *Proteomics*, 4(6):1551–1561, 2004. Publisher: Wiley Online Library.

[49] Sarah Hunter, Rolf Apweiler, Teresa K Attwood, Amos Bairoch, Alex Bateman, David Binns, Peer Bork, Ujjwal Das, Louise Daugherty, Lauranne Duquenne, and others. InterPro: the integrative protein signature database. *Nucleic acids research*, 37(suppl_1):D211–D215, 2009. Publisher: Oxford University Press.

[50] Min-Seok Kwon, Sang Yun Cho, and Young-Ki Paik. Protein databases. In *Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine*, pages 1483–1487. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.

[51] Minoru Kanehisa, Miho Furumichi, Mao Tanabe, Yoko Sato, and Kanae Morishima. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*, 45(D1):D353–D361, January 2017.

[52] Michael Ashburner, Catherine A Ball, Judith A Blake, David Botstein, Heather Butler, J Michael Cherry, Allan P Davis, Kara Dolinski, Selina S Dwight, Janan T Eppig, and others. Gene ontology: tool for the unification of biology. *Nature genetics*, 25(1):25–29, 2000. Publisher: Nature Publishing Group.

[53] The Gene Ontology Consortium, Seth Carbon, Eric Douglass, Benjamin M Good, Deepak R Unni, Nomi L Harris, Christopher J Mungall, Basu, et al. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Research*, 49(D1):D325–D334, January 2021.

[54] Brett Trost and Anthony Kusalik. Computational phosphorylation site prediction in plants using random forests and organism-specific instance weights. *Bioinformatics*, 29(6):686–694, January 2013. _eprint: https://academic.oup.com/bioinformatics/article-pdf/29/6/686/17103708/btt031.pdf.

[55] Christiam Camacho, George Coulouris, Vahram Avagyan, Ning Ma, Jason Papadopoulos, Kevin Bealer, and Thomas L Madden. BLAST+: architecture and applications. *BMC Bioinformatics*, 10(1):421, December 2009.

[56] Shakiba Jalal, Ryan Arsenault, Andrew A Potter, Lorne A Babiuk, Philip J Griebel, and Scott Napper. Genome to kinome: species-specific peptide arrays for kinome analysis. *Science signaling*, 2(54):pl1–pl1, 2009.

[57] Conor Lazarou. Peptide microarray quality control, May 2018.

[58] Adán José-García and Wilfrido Gómez-Flores. A survey of cluster validity indices for automatic data clustering using differential evolution. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 314–322, 2021.

[59] William M. Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(336):846–850, December 1971.

[60] Peter J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65, 1987.

[61] Mahesh Visvanathan, B Srinivas Adagarla, H Lushington Gerald, and Peter Smith. Cluster validation: An integrative method for cluster analysis. In *2009 IEEE International Conference on Bioinformatics and Biomedicine Workshop*, pages 238–242, 2009.

[62] Nadia Bolshakova, Francisco Azuaje, and Pádraig Cunningham. An integrated tool for microarray data clustering and cluster validity assessment. *Bioinformatics*, 21(4):451–455, 12 2004.

[63] Miljana Tanić, Kira Yanowski, Eduardo Andrés, Gonzalo Gómez-López, María Rodríguez-Pinilla Socorro, David G Pisano, Beatriz Martinez-Delgado, and Javier Benítez. mirna expression profiling of formalin-fixed paraffin-embedded (ffpe) hereditary breast tumors. *Genomics data*, 3:75–79, 2015.

[64] Susan K Munster, Vicky L White, David C Hutchings, Dennis M Burian, Scott J Nicholson, LLC Venesco, et al. Comparison study of microarray and rna-seq for differential expression. Technical report, United States. Department of Transportation. Federal Aviation Administration . . . , 2018.

[65] R. Core Team. R: A language and environment for statistical computing. 2021. R Foundation for Statistical Computing, Vienna, Austria.

[66] Larry Wall. The Perl programming language, 1994.

[67] Ben Laurie and Peter Laurie. *Apache: The definitive guide.* " O'Reilly Media, Inc.", 2003.

[68] Audrey Kauffmann, Robert Gentleman, and Wolfgang Huber. arrayQualityMetrics—a bioconductor package for quality assessment of microarray data. *Bioinformatics*, 25(3):415–416, 2009. Publisher: Oxford University Press.

[69] Guido van Rossum. Python reference manual. *Department of Computer Science [CS]*, (R 9525), 1995. Publisher: CWI.

[70] John D Hunter. Matplotlib: A 2d graphics environment. *Computing in science & engineering*, 9(03):90–95, 2007.

[71] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020. Publisher: Springer Science and Business Media LLC.

[72] Michael L. Waskom. seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60):3021, 2021. Publisher: The Open Journal.

[73] Oliver Kramer. Scikit-learn. In *Machine learning for evolution strategies*, pages 45–53. Springer, 2016.

[74] Shizuo Akira and Kiyoshi Takeda. Toll-like receptor signalling. *Nature Reviews Immunology*, 4(7):499–511, July 2004.

[75] Sander H. Diks, Klaartje Kok, Tom O'Toole, Daan W. Hommes, Peter van Dijken, Jos Joore, and Maikel P. Peppelenbosch. Kinome profiling for studying lipopolysaccharide signal transduction in human peripheral blood mononuclear cells. *The Journal of Biological Chemistry*, 279(47):49206–49213, November 2004.

[76] Ryan J. Arsenault, Shakiba Jalal, Lorne A. Babiuk, Andrew Potter, Philip J. Griebel, and Scott Napper. Kinome analysis of Toll-like receptor signaling in bovine monocytes. *Journal of Receptors and Signal Transduction*, 29(6):299–311, December 2009.

[77] Werner Feist, Artur J Ulmer, Ming-Hai Wang, Joachim Musehold, Carsten Schlüter, Johannes Gerdes, Hildegard Herzbeck, Helmut Brade, Shoichi Kusumoto, Tibor Diamantstein, et al. Modulation of lipopolysaccharide-induced production of tumor necrosis factor, interleukin 1, and interleukin 6 by synthetic precursor ia of lipid a. *FEMS Microbiology Immunology*, 4(2):73–89, 1992.

[78] Junsheng Dong, Jianji Li, Luying Cui, Yefan Wang, Jiaqi Lin, Yang Qu, and Heng Wang. Cortisol modulates inflammatory responses in LPS-stimulated RAW264.7 cells via the NF-kB and MAPK pathways. *BMC Veterinary Research*, 14(1):30, December 2018.

[79] NM Foley, JM Racz, Z Al-Hilli, V Livingstone, T Cil, CMB Holloway, L Romics, Z Matrai, MW Bennett, L Duddy, et al. An international multicenter review of the malignancy rate of excised papillomatous breast lesions. *Annals of surgical oncology*, 22(3):385–390, 2015.

[80] Tamara T Laha, Marguerite Hawley, Kenneth L Rock, and Alfred L Goldberg. Gamma-interferon causes a selective induction of the lysosomal proteases, cathepsins b and l, in macrophages. *FEBS letters*, 363(1-2):85–89, 1995.

[81] Antonio Facciuolo, Patricia Gonzalez-Cano, Scott Napper, Philip J Griebel, and Lucy M Mutharia. Marked differences in mucosal immune responses induced in ileal versus jejunal Peyer's patches to Mycobacterium avium subsp. paratuberculosis secreted proteins following targeted enteric infection in young calves. *PLoS One*, 11(7):e0158747, 2016. Publisher: Public Library of Science San Francisco, CA USA.

[82] Boris Iglewicz and David Hoaglin. Volume 16: how to detect and handle outliers. *The ASQC basic references in quality control: statistical techniques*, 16, 1993.

[83] Lars M. T. Eijssen, Magali Jaillard, Michiel E. Adriaens, Stan Gaj, Philip J. de Groot, Michael Müller, and Chris T. Evelo. User-friendly solutions for microarray quality control and pre-processing on ArrayAnalysis.org. *Nucleic Acids Research*, 41(W1):W71–W76, July 2013.

[84] Audrey Kauffmann and Wolfgang Huber. Microarray data quality control improves the detection of differentially expressed genes. *Genomics*, 95(3):138–142, March 2010.

[85] Andreas Buness, Wolfgang Huber, Klaus Steiner, Holger Sültmann, and Annemarie Poustka. arraymagic: two-colour cdna microarray quality control and preprocessing. *Bioinformatics*, 21(4):554–556, 2005.

[86] Amnon Koren, Itay Tirosh, and Naama Barkai. Autocorrelation analysis reveals widespread spatial biases in microarray experiments. *BMC Genomics*, 8(1):164, 2007.

[87] Ofer Shai, Q. Morris, and Brendan J. Frey. Spatial bias removal in microarray images. *University of Toronto Technical Report PSI-2003-21*, 2003.

[88] Alex M. Dussaq, Timothy Kennell Jr, Nicholas J. Eustace, Joshua C. Anderson, Jonas S. Almeida, and Christopher D. Willey. Kinomics toolbox—A web platform for analysis and viewing of kinomic peptide array data. *PLOS ONE*, 13(8):e0202139, August 2018.

[89] Christoph Endrullat, Jörn Glökler, Philipp Franke, and Marcus Frohme. Standardization and quality management in next-generation sequencing. *Applied & translational genomics*, 10:2–9, 2016.

[90] Ankita Negi, Abhimati Shukla, Akanksha Jaiswar, Jatin Shrinet, and Rahul Singh Jasrotia. Applications and challenges of microarray and rna-sequencing. *Bioinformatics*, pages 91–103, 2022.