A Novel Feature Selection Approach for Data Integration Analysis: Applications to Transcriptomics study

by

Nisha Puthiyedth

M.Sc. & B.Tech. in Bioinformatics

Thesis

submitted in partial fulfilment

of the requirements for the Degree of

Doctor of Philosophy

THE UNIVERSITY OF
NEWCASTLE
AUSTRALIA

School of Electrical Engineering
and Computer Science

November, 2016

This reaserach work was performed under the supervision of Prof. Pablo Moscato, A/Prof. Regina Berretta and Dr. Carlos Riveros.

Statement of Authorship and Collaboration

I hereby certify that the work embodied in this thesis contains two published papers of which I am the first author. Part of the paper titled "A new combinatorial optimization approach for integrated feature selection using different datasets: a prostate cancer transcriptomic study" published on PLOS One (Nisha Puthiyedth, Carlos Riveros, Regina Berretta and Pablo Moscato, 2015, [284]) is given in Chapter 4. Part from the paper titled "Identification of Differentially Expressed Genes Through Integrated Study of Alzheimer's Disease Affected Brain Regions" which is published in PLOS One (Nisha Puthiyedth, Carlos Riveros, Regina Berretta and Pablo Moscato, 2016 [285]) is given in Chapter 5. Nevertheless, it is worth mentioning that I had an active role in every stage of the development of these papers. I include this written statement, endorsed by my supervisors, attesting to my contribution to the aforesaid papers.

Nisha Puthiyedth
9th November 2016

Dr. Carlos Riveros
9th November 2016

A/Prof. Regina Berretta
9th November 2016

Prof. Pablo Moscato
9th November 2016

This thesis is dedicated to my husband, son, brother, mother and father for their endless love, care and encouragement.

Acknowledgements

This thesis represents not only my work at the keyboard but also the result of many experiences I have encountered at the Centre for Bioinformatics, Biomarker Discovery and Information-Based Medicine (CIBM) from dozens of wonderful individuals.

First and foremost I wish to thank my supervisor, Professor Pablo Moscato, director of CIBM for his continuous support throughout my PhD and related research with patience, motivation and immense knowledge. I would also like to thank my co-supervisors, A/Prof Regina Berretta and Dr Carlos Riveros. Regina, a very creative and disciplined person, helped me to be on track throughout these three and half years. Her guidance helped me in all the time of research, especially in the writing phase of research. Carlos, very enthusiastic person, one of the best persons I have seen for helping and understanding others. I have been extremely lucky to have a co-supervisor like him who cared so much about my work, and responded to my questions and queries so promptly. I thank Carlos not only for supervising me, but also for so many other things he helped me with. It is always been a pleasure to work with someone like him who is with open mind, great energy and always a smile on his face. Thanks Carlos for teaching me loads of things that I can carry for the rest of my journey.

I also thank the whole CIBM team; Ademir Gabardo, Ahmed Arefin, Amer Abu Zaher, Amir Salehipour, Claudio Sanhueza , Francia Jimenez, Heloisa Zaccaron, Inna Tishchenko, Leila Moslemi Naeni, Luke Mathieson, Mohammad Nazmul Haque, Nasimul Noman, Natalie De Vries, Renato Vimieiro, Shannon Fenn ; for the discussion, fun and the moments we shared. Special thanks go to Nasimul Noman for his support and advice.

I would like to thank my family, husband, mother, son and brother, for supporting me throughout my life, financially, practically and with moral support. The support they have given is unconditional in all these years. Especially my husband, Shajil, have given up many things to support my studies. Also I thank my friends for their support and encouragement.

I am very thankful to the University of Newcastle for providing me an opportunity to be a part of the research environment with all facilities and scholarships. My PhD was funded by the University of Newcastle Postgraduate Research Scholarship and National Health and Medical Research Council (NHMRC) grant.

Finally, I would like to thank the Elite editing for providing me with the proof reading service and the University of Newcastle for providing the grant for the same. Also I thank the reviewers for thier valuable feedback and advise.

To all the others that are not mentioned here, please forgive me and be sure that you were also important in my journey.

Contents

List of Tables . . . . . **xiv**

List of Figures . . . . . **xvi**

1 Introduction **1**

1.1 Motivation for the study 1

1.2 Meta-analysis: an overview 3

1.2.1 Observational selection of evidence 4

1.2.2 Standardised outcome measure 4

1.2.3 Calculate the overall effect 4

1.2.4 Sensitivity analysis 5

1.3 Goals 6

1.4 Thesis Overview 7

2 Meta-analysis and Microarray Data Research **9**

2.1 Meta-analysis methods for microarray studies 10

2.1.1 Microarray technology 10

2.1.2 Meta-analysis of microarray studies 13

2.2 Conclusion 17

3 Proposed Methods for Feature Selection **19**

3.1 Feature selection approaches 19

3.1.1 Unsupervised methods 20

3.1.2 Supervised Methods 20

3.2 The $(\alpha, \beta) - k$ Feature Set Problem approach 22

3.2.1 The problem definition 23

3.2.2 The graph representation 25

3.2.3 Previous studies using the (α, β)-k Feature Set problem approach 27

3.3 The Coloured (α, β)-k Feature Set problem approach 29

3.3.1 The problem definition 29

3.3.2 The graph representation . . . 33
3.4 The Generalised (α, β)-k Feature Set problem approach . . . 35
3.4.1 The problem definition . . . 35
3.4.2 The graph representation . . . 39
3.5 Mathematical formulation . . . 39
3.6 Conclusion . . . 45

4 Application: Prostate Cancer datasets . . . 47
4.1 Cancer . . . 48
4.1.1 Prostate Cancer . . . 49
4.2 Datasets . . . 50
4.3 Individual dataset analysis . . . 52
4.3.1 Application of the (α, β)-k Feature Set problem approach 53
4.3.2 Application of t-test . . . 54
4.4 Integrated data analysis . . . 56
4.4.1 Integration method . . . 56
4.4.2 Application of the Coloured (α, β)-k Feature Set problem approach . . . 59
4.5 Comparison . . . 62
4.6 Sensitivity analysis . . . 63
4.7 Functional and Pathway Analysis . . . 63
4.8 Discussion . . . 67
4.9 Conclusion . . . 69

5 Application: Alzheimer's Disease Brain Regions . . . 71
5.1 Alzheimer's Disease . . . 71
5.2 Datasets . . . 75
5.2.1 Individual Data Analysis . . . 76
5.3 Integrated data analysis . . . 94
5.3.1 Integration method . . . 94
5.3.2 Application of the Coloured (α, β)-k Feature Set problem approach . . . 94
5.4 Comparison . . . 100
5.5 Sensitivity Analysis . . . 101
5.6 Discussion . . . 102
5.7 Conclusion . . . 112

6 Conclusions and Future work . . . 113
6.1 Conclusion . . . 113
6.2 Future Work . . . 116

6.3 Final Notes . . . 118
6.4 Data Sources . . . 119

7 Appendix A . . . 121
7.1 Glossary . . . 121

8 Appendix B . . . 123
8.1 List of results . . . 123
8.1.1 Chapter 2 - Meta-analysis methods . . . 123
8.1.2 Chapter 3 - Results . . . 129
8.1.3 Chapter 4 - Results . . . 132
8.1.4 Chapter 5 - Results . . . 147

9 Appendix C . . . 231
9.1 R source code used for the data analysis . . . 231

List of Tables

2.1 Example of a gene expression matrix. 12

3.1 An example of a numerical data with seven features and five samples. 24
3.2 Transformed data into a coverage 25
3.3 An example of numerical data for blue dataset (hence the suffix 'b' in the sample names) with five features and five samples. . . 31
3.4 An example of numerical data for an orange dataset (hence the suffix 'o' in the sample names) with five features and five samples. 31
3.5 Transformed value to a coverage for blue and orange datasets. . 32
3.6 An example of numerical data for *first* dataset with five features and five samples. 36
3.7 An example of a numerical data for *second* dataset with five features and five samples. 37
3.8 Transformed value to a coverage for datasets f (first) and s (second). 38

4.1 Summary of prostate cancer datasets used in this study 51
4.2 The results of the numerical solution to the (α, β)-k Feature Set problem approach for each of the six individual datasets. 53
4.3 List of common genes among all the individual dataset results from Table 4.2 54
4.4 t-test results on individual dataset. 55
4.5 List of common genes among all the individual dataset results of the (α, β)-k Feature Set problem approach from Table 4.4 . . 55
4.6 The details for the probes in dataset 1. 58
4.7 The details for the probes in Dataset 2. 58
4.8 Result of the Coloured (α, β)-k Feature Set problem approach . 59
4.9 Overlapping genes in t-test, the (α, β)-k Feature Set problem approach and the Coloured (α, β)-k Feature Set problem approach. 61

4.10 Comparison of the Coloured (α, β)-k Feature Set problem approach and RankProd result 62
4.11 Result of sensitivity analysis. 64
4.12 The top 14 resulted pathways from pathway analysis. 66

5.1 Sample details that belongs to different regions. 76
5.2 Significant genes in different brain regions. 78
5.3 List of dysregulated pathways related with EC region 79
5.4 List of microRNAs resulted from EC region. 80
5.5 List of dysregulated pathways related with HIP region 81
5.6 List of microRNAs resulted from HIP region. 83
5.7 List of dysregulated pathways related with MTG region 85
5.8 List of microRNAs resulted from MTG region. 88
5.9 List of dysregulated pathways related with PC region 89
5.10 List of microRNAs resulted from PC region. 90
5.11 List of dysregulated pathways related with SFG region 91
5.12 List of microRNAs resulted from SFG region. 92
5.13 List of microRNAs resulted from VCX region. 93
5.14 List of non-coding features resulted from the Coloured (α, β)-k Feature Set and Generalised (α, β)-k Feature Set problem approach. 98
5.15 Dysregulated pathway related to non-coding RNAs. 99
5.16 Top dysregulated pathways 103
5.17 The top most DEGs across EC, HIP, MTG, PC and SFG. . . . 105

List of Figures

3.1 Graph representation of the (α, β)-k Feature Set problem instance created from Table 3.2. Blue nodes represent sample pairs that belongs to different classes (α node), yellow nodes represent the feature node and white nodes represent sample pairs that belong to same class (β node). The figure adapted from [20] 26

3.2 An optimal solution for the (α, β)-k Feature Set problem instance from Figure 3.1 with $\alpha = 1$ and $\beta = 1$. The figure was adapted from [20] 27

3.3 Graph representation of the Coloured (α, β)-k Feature Set problem created from Table 3.5. Each blue nodes represents sample pairs that belong to the 'blue' dataset, *'b'*; each orange node represents sample pairs that belong to the 'orange' dataset, *'o'* and each yellow node represents the features. 34

3.4 An optimal solution for the Coloured (α, β)-k Feature Set problem instance from Figure 3.3 with $\alpha = 1$ and $\beta = 1$. 35

3.5 Graph representation of the generalised (α, β)-k Feature Set problem created from Table 3.8. Blue nodes represent sample pairs that belong to different classes, α nodes; white nodes represent sample pairs that belong to the same class, β nodes and yellow nodes represent the features. 40

3.6 An optimal solution for the generalised (α, β)-k Feature Set problem instance from Figure 3.5 with $\alpha = 1$ and $\beta = 1$. 41

4.1 Cancer stem cells undergoing excessive symmetric cell division. **(A)** A normal cell maintains a controlled series of symmetric and asymmetric cell divisions.**(B)** A cell undergoes excessive symmetric cell divisions due to dysregulated self-renewal, and forms a tumour. *The figure is adapted from [282] with permission.* 48

4.2 Prostate cancer development; showing different stages of prostate cancer. *The figure is adapted from [67]with permission* . . 50

4.3 Heatmap for the Coloured (α, β)-k Feature Set problem approach resulted genes that cover five or more datasets. It contains 186 up and down regulated genes (rows). The blocks of greenish blue colour represent the absence of gene values in particular datasets. The first colour bar at the bottom indicates Primary Tumour (blue) and Normal (green) samples. The second colour bar represents each sample group in different colour. L-2695 (blue), L-3044 (red), L-3289 (orange), Welsh (grey), Uma (cyan) and Singh (dark grey).The row names of the heatmap is given in Appendix 8.1.3.The magnified version of the figure is available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4480358/figure/pone.0127702.g001/ . . . 60

4.4 Heatmap for the Coloured (α, β)-k Feature Set problem approach resulted 120 genes that cover six datasets. There are 120 up and down regulated genes (rows) which are differentially expressed between normal and tumour classes. The two colour bars at the bottom represent the ordering of samples and sample groups, respectively, as explained in Figure 4.3.The row names of the heatmap is given in Appendix 8.1.3. The magnified version of the figure is available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4480358/figure/pone.0127702.g002/ 61

5.1 Heatmap for the 479 features with BF-value < 0.0001 in the brain region-combined analysis. There were 479 up- and down-regulated genes that were differentially expressed between control and AD samples. The first colour bar at the bottom indicates AD (blue) and control (red) samples. The second colour bar represents each sample group in different colours: EC (blue), HIP (red), MTG (orange), PC (grey) and SFG (cyan). The magnified version of the figure is available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4822961/figure/pone.0152342.g001/ 95

5.2 Heatmap for the 540 features with BF-value < 0.0001 from the combined analysis. There were 540 up- and down-regulated probes that were differentially expressed between control and AD sample. The first colour bar at the bottom indicates AD (blue) and control (green) samples. The second colour bar represents each sample group in a different colour: EC (blue), HIP (red), MTG (orange), PC (grey) and SFG (cyan). The magnified version of the figure is available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4822961/figure/pone.0152342.g002/ 97

5.3 Heatmap for the 23 non-coding features resulting from the combined analysis. These 23 up- and down-regulated features were differentially expressed between control and AD samples. The first colour bar at the bottom indicates AD (blue) and control (red) samples. The second colour bar represents each sample group in a different colour: EC (blue), HIP (red), MTG (orange), PC (grey) and SFG (cyan). The magnified version of the figure is available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4822961/figure/pone.0152342.g003/ 100

Abstract

Meta-analysis has become a popular method for identifying novel biomarkers in the field of medical research. Meta-analysis has been widely applied to genome-wide association and transcriptomic studies due to the availability of datasets in the public domain. Joint analysis of multiple datasets has become a common technique for increasing statistical power in detecting biomarkers reported in smaller studies. The approach generally followed relies on the fact that as the total number of samples increases, greater power to detect associations of interest is anticipated. Integrating available information from different datasets to generate a combined result seems reasonable and promising. Consequently, there is a need for computationally based integration methods that evaluate multiple independent datasets investigating a common theme or disorder. This raises a variety of issues in the analysis of such data and leads to more complications than are seen with standard meta-analysis, including diverse experimental platforms and complex data structures. I illustrate these ideas using microarray datasets from multiple studies and propose an integrative methodology to combine datasets generated using different platforms. Having combined the data, the main challenge is to choose a subset of features that represent the combined dataset in a particular aspect. While the approach is well established in biostatistics, the introduction of new combinatorial optimisation models to address this issue has not been explored in depth.

In 2004, a new feature selection approach based on a combinatorial optimisation method was proposed, entitled the (α, β)-k Feature Set problem approach. The main advantage of this approach over ranking methods for selecting individual features is that the features are evaluated as groups instead of on the basis of their individual performance. The (α, β)-k Feature Set problem approach has been defined having first in mind a single uniform dataset, and conceived in this ways, it is not readily applicable to the case of integrated datasets. An extended version of this approach handles integrated datasets in a consistent manner and selects features that differentiate sample pairs across datasets. The application of an (α, β)-k Feature Set problem -based approach

for meta-analysis thus helps to identify the best set of features from a combined dataset, allowing researchers to reveal the genetic pathways that contribute to the development of a disease. I propose an extended version of the (α, β)-k Feature Set problem approach that aims to find a set of genes whose expression level may be used to identify a joint core subset of genes that putatively play an important role in two conditions: prostate cancer and Alzheimer's disease.

The results of the current study suggest that the proposed method is an efficient meta-analysis method that is capable of identifying biologically relevant genes that other methods fail to identify. As the amount of data increases, this novel method can be applied to find additional genes and pathways that are significant in these diseases, which may provide new insights into the disease mechanism and contribute towards understanding, prevention and cures.

Chapter 1

Introduction

This chapter gives an overview of the thesis with the motivation for this study in Section 1.1 and an introduction, procedure and principles for meta-analysis in Section 1.2. The goals of the study are explained in Section 1.3. The structure of the thesis and details about the content of each chapter are given in Section 1.4.

1.1 Motivation for the study

The extraction of information arising from the integration of multiple datasets and its translation into domain knowledge is a significant problem in several fields. Biology and health studies globally are increasingly applying the useful policy of making their raw data and results available for the common good via public domain databases. This open sharing has enhanced the reproducibility of other researchers' findings and enabled the general scrutiny of results by the wider scientific community. Existing online datasets are also becoming very useful for the development of new mathematical and computational approaches for pattern recognition, machine learning and artificial intelligence methods.

The healthy practice of sharing data is now being increasingly adopted by governments and scientific journals. The private and public sector is also engaging in 'data-mining competitions' in which datasets are made widely available and crowd-sourced for data analysis. This new, digital and interconnected, global research, open data enterprise is undoubtedly a positive direction for science, research and development and it can be confidently stated that this trend is here to stay.

The existence of a large number of publicly available datasets provides strong motivation for the development of new mathematical methods that help to extract panels of biomarkers by employing several datasets. However,

despite the growing number of studies, overall consensus has yet to be reached about how to do this [111].

Researchers often only highlight the obstacles ahead–for instance, by pointing at the essential differences among microarray platforms, experimental designs, collection procedures for samples, heterogeneity of laboratory protocols and analysis methods used for a study [187]. Most biological studies are unable to provide a definitive answer to the question of interest because of low sample sizes [69]. All of these confounding issues must be considered and highlighting them does not diminish the need to develop integrative techniques for a joint panel for biomarker elicitation.

Many biological studies have shown that it is difficult to obtain a reliable result from a single dataset [38, 256, 294, 389]. Although some researchers may eventually procure the financial resources to conduct studies with a large number of samples – leading to greater power to detect individual markers – an integrated study can provide a clearer picture as the final result would identify consensus among a number of individual studies. This demonstrates the need for new approaches to construct a list of significant features (in the current context, genes) from multiple platforms when looking at a panel that acts together to complete a discrimination task across several studies.

In 1988, Huque [150] defined meta-analysis as 'a statistical analysis that combines or integrates the results of several independent clinical trials considered by the analyst to be "combinable" '. Meta-analysis helps to obtain a clear and solid conclusion for a particular hypothesis using a more appropriate methodological approach. When some conditions are met, an integrated study can help to improve the power of the analysis by increasing the total number of samples under consideration [117]. Meta-analysis is also an important and effective approach when existing studies have conflicting conclusions [139] and the overall aim is to resolve them if possible.

Increasing the detection power of smaller studies by integrating them in a larger study has also become a way to overcome research funding limitations. This is particularly true for transcriptomics, a field in which there is an undeniable need for new mathematical models and algorithms aimed at extracting information, by jointly studying multiple datasets from different and ever-changing technological platforms.

Multi-platform data integration remains challenging as datasets from different experiments are not directly comparable due to factors associated with the generation of the dataset [153]. Some of the challenges are simply technical in nature: for instance, genomic data can be collected in a wide variety of data formats, thus making direct integration difficult. Datasets can be converted to

a common data format before combining them, but this is not always feasible [125]. A new method for integrating datasets without the need to transform values to a common uniform format and range can help to overcome these problems.

1.2 Meta-analysis: an overview

Meta-analysis can be defined as the quantitative review and statistical evaluation of a collection of studies that are related but independent, for the purpose of integrating the findings [265]. In 1904, Karl Pearson [170] introduced meta-analysis, a method that combines the results of different studies to generate more powerful statistics than would be provided by analysing individual studies. In 1940, Pratt et al. [283] conducted a meta-analysis of the results of identical experiments concerning a particular research issue. Since then, meta-analysis has been widely used in clinical trials and epidemiological studies [217, 71, 341, 338, 343, 122, 375], especially those employing microarray datasets [290, 295, 294, 41, 42, 394, 47, 50]. Indeed meta-analyses are the most frequently cited studies in clinical research [264]. Meta-analysis is mainly used to resolve contradictions found when analysing the results of different studies, in the quest to translate the joint findings from those studies.

Randomised clinical trials have been used for over 70 years to test the efficacy of disease conditions in a research setting with highly selected participants under controlled conditions. However, clinical trials are very costly, which increases the value of meta-analysis to retrieve more information from existing datasets. When applied correctly, meta-analysis is a highly efficient way to obtain different endpoints by using suitable statistical techniques [244]. Moreover, the expansion of public data repositories creates a need for meta-analysis to evaluate, integrate and validate the related datasets.

A key component of meta-analysis is the systematic review of a focussed question through an extensive search. Two approaches used for this purpose in meta-analysis are literature- or summary-based searches and patient data-based searches.

1. The literature or summary based search
   Literature- or summary-based meta-analysis is the analysis of the results from existing studies to obtain a balanced conclusion by combining those results [62].

2. The patient data based search
   Patient data-based meta-analysis is the most useful and effective approach, in which the original data from each study are collected and integrated for further analysis [62].

In both approaches, meta-analysis should be carefully planned with a detailed written protocol and procedure. A typical plan for meta-analysis is discussed in the following sections.

1.2.1 Observational selection of evidence

Eligibility criteria for inclusion of a study in a meta-analysis must be defined through a comprehensive search of studies. Selection should be based mainly on the quality and type of trial, the patients, outcomes and the lengths of follow-up. However, the quality and type of trial is highly dependent on the choices of the researcher and hence the final result can be subjective [304].

Strategies for selecting relevant studies depend on the research question at hand. Also, a decision must be made as to whether to include only published studies or to extend the search to unpublished studies. In a meta-analysis, it is worth including unpublished studies as they may contain more information than published ones.

A published study search can be performed using databases such as PubMed [301]. After selecting the studies to be used in the analysis, the relevant datasets can be downloaded in a standardised form.

1.2.2 Standardised outcome measure

Individual study results must be standardised to a homogeneous form to enable comparison between them. If the end point measure is continuous (e.g. blood pressure), the standard deviation between the treatment and control groups is used to have a comparison benchmark. If the end point measure is binary (e.g. disease *v.* controls), odds ratios or relative risks are used. Using odds ratios makes it easy to combine data and test the overall effect in the case of binary end points when comparing the data with continuous end points data.

1.2.3 Calculate the overall effect

The overall effect can be calculated by combining all selected datasets. Statistical models used to analyse the combined data can be broadly classified as either fixed effects or random effects (I refer the reader to Appendix 8.1.1 for

more details). The choice of method will depend on the way in which the data are treated and the variability among individual study results.

1.2.4 Sensitivity analysis

The robustness of findings and the quality of the selected or proposed methodology can be examined through a sensitivity analysis. The most common way of performing a sensitivity analysis is via the 'leave-one-out' approach [334] or by repeating the analysis using different methods and comparing the results between them.

A carefully constructed and implemented meta-analysis that follows these procedures can help scientists to develop or confirm a specified theory about a set of phenomena under investigation. However, like every other method, meta-analysis has strengths and weaknesses:

- they can help to summarise the findings of different studies [352].
- they can reduce the uncertainty of interpretations [223].
- the heterogeneity in study results can be examined [68].
- they can increase precision of literature reviews [12].
- they have larger sample sizes than individual studies [68].
- they can suggest research questions for future studies [352].
- the quality of a study can be improved [68].
- they can perform more objective assessment of evidence and reduce disagreement [12].
- they can help to introduce effective treatments into clinical practice [95].
- they can reveal the flaws and biases in the procedures and results of a single study [321].
- a good meta-analysis of badly designed studies will still result in bad statistics [68].
- they have a high chance of showing publication bias [12].
- the methodology used for a meta-analysis has a substantial effect on the final results [223].

- failing to account for all the relevant factors in the methodology may lead to inaccurate conclusions [223].

Despite the substantial advantages of meta-analyses over individual studies, there are some important factors that affect a particular meta-analysis, such as the association between the selected studies and the heterogeneous nature of the samples it uses. The results of a meta-analysis are only likely to be reliable if it is based on high-quality studies and methodologies.

With the growth in high-quality clinical studies that use recent methodologies, appropriate datasets will become increasingly available for researchers. To maximise the advantages offered by the availability of these datasets, it is essential to develop new methods for meta-analysis.

1.3 Goals

Advances in technology promote a dramatic increase in data availability in different domains. In the field of biology, particularly bioinformatics, technologies such as microarrays and next-generation sequencing have delivered enormous amounts of data. Such data can significantly increase computational difficulties in handling them, so improved analytical approaches are urgently required.

Many methods are available for the individual analysis of these datasets, but not for the integrated analysis of high-dimensionality datasets. The first aim of this study is to propose an integration method to combine datasets generated using different platforms. Even if datasets are integrated in an efficient manner, selecting significant features from the multi-platform integrated dataset is a challenging topic in the field of bioinformatics. Therefore, this thesis focuses on the development of a new feature selection method that can deal with integrated datasets, and particularly with the analysis and identification of differentially expressed genes (DEGs) that belong to different groups.

Strictly following the procedures explained in Section 1.2, I perform a meta-analysis of microarray datasets generated using different platforms as an application of the proposed method. I perform the meta-analysis of these datasets and provide a comparison of the method with several popular, state-of-the-art meta-analysis methodologies. I also perform and report on a sensitivity analysis to check the robustness of the findings.

1.4 Thesis Overview

This thesis is structured as six chapters: Introduction; Meta-analysis and Microarray Data Research; Proposed Method for Feature Selection; Application of the Coloured (α, β)-k Feature Set Problem Approach to Prostate Cancer Datasets; Application of the Coloured (α, β)-k Feature Set Problem Approach to Alzheimer's Disease (AD) Datasets; and Conclusions and Future works.

This chapter, Chapter 1 - Introduction has provided an overview of meta-analysis and the motivation and goals of this study.

Chapter 2 - Meta-analysis and Microarray Data Research gives an detailed overview of meta-analysis methods. As microarray datasets are used for the application of the proposed method, microarray technology is also introduced in this chapter. The introduction to microarray technologies provides a detailed picture of the data generated using different platforms and how their quality is assessed. A literature review on the meta-analysis of microarray studies is then provided and the methods used for comparison with the proposed method are explained in detail.

Chapter 3 - Proposed Methods for Feature Selection- provides a brief explanation of feature selection approaches and introduces the proposed feature selection method, the Coloured (α, β)-k Feature Set problem approach, which is a variant of the (α, β)-k Feature Set problem approach. The latter is explained in detail along with its previous applications in other problem areas. The decision version, along with the integer programming model of the Coloured (α, β)-k Feature Set problem approach, is also presented.

Chapter 4 - Application: Prostate Cancer datasets - begins with a brief explanation of prostate cancer and details about the prostate cancer datasets selected for the current study. The individual analysis of the selected datasets and their results are explained with the application of the (α, β)-k Feature Set problem approach and t-test [230]. The proposed integration method for combining datasets is then explained in detail. The application of the Coloured (α, β)-k Feature Set problem approach to the integrated dataset and the resulting signatures are reported in this chapter. The results of functional and pathway analysis of the resulting DEGs are provided. The sensitivity analysis results and a comparison of the Coloured (α, β)-k Feature Set problem approach results with those from popular meta-analysis methods, RankProd and GeneMeta, are also presented. Finally, validation of the DEGs in terms of prostate cancer development is performed through a literature search and is presented.

Chapter 5- Application: Alzheimer's Disease Brain Regions - provides an introduction to (AD) and the datasets selected for the current study. The integration process for the datasets and the application of the Coloured (α, β)-k Feature Set problem approach is explained, along with the results. A sensitivity analysis to evaluate the robustness of the proposed method is provided. The results are then compared with those from RankProd and GeneMeta.

Chapter 6- Conclusions and Future work concludes this thesis by summarising the key findings from Chapter 4, Chapter 5.

Chapter 2

Meta-analysis and Microarray Data Research

In this chapter, a brief explanation of microarray technology is provided in Section 2.1 as microarray datasets are used for the application of the proposed method. A review of the literature on meta-analysis using microarray datasets is presented in Subsection 2.1.1. The meta-analysis methods used here for comparative analyses are described in Subsection 2.1.2. Finally, a conclusion for the chapter is presented in Section 2.2.

Meta-analysis is one of the most complex and powerful research tools available to combine different studies or datasets and create a single study for analysis, in an effective manner. Since 1904, meta-analysis has been widely used in scientific research. Being a fundamental science on its own, meta-analysis is mostly used in healthcare sciences to integrate available biological information and assist with generalisation of findings. The literature identifies two major categories of meta-analysis: one tests the statistical significance of combined results and the other combines the effect size of individual studies. These are the most known meta-analysis methods, but are out of the scope this thesis, hence the popular methods in each category, and their application, are discussed in Appendix 8.1.1.

In this study we use microarray datasets for the application of our proposed method. So we concentrated on the meta-analysis of microarray studies, which are explained in the following sections.

2.1 Meta-analysis methods for microarray studies

The existence and increasing amount of publicly available microarray study results provides a strong motivation to perform meta-analysis of those studies in an attempt to extract panels of biomarkers. The greatest challenge is the efficient integration and feature selection of microarray data generated by different research groups on different array platforms. Microarray technology and data assessment are explained in the following sections.

2.1.1 Microarray technology

Deoxyribonucleic acid (DNA) microarrays depend on the hybridisation of complementary nucleic acid sequences of DNA fragments called probes, which are immobilised on a solid surface (glass slide, silicon chip or nylon membrane) in a high-density arrangement [332]. The target, unknown genomic messenger ribonucleic acid (mRNA) sequence of interest, is then fluorescently labelled and hybridised to the probe. Successful hybridisation between the labelled target and immobilised probe will result in increased fluorescence over a background level, which can be measured using a fluorescent scanner. The fluorescence data captured by the scanner can then be analysed using a range of methods [240].

DNA microarrays can be divided into two major types based on the immobilised probes used to build the array: complementary DNA (cDNA) arrays and oligonucleotide arrays. In cDNA arrays, the DNA fragments are generated using polymerase chain reaction (PCR) and may vary greatly in length. Nucleic acids with any length, composition or origin can be arrayed for cDNA arrays. Hence the specificity of hybridisation can be greater than with oligonucleotide arrays. Prior knowledge about the sequence is not required as clones from a cDNA library can be used and then sequenced, depending on the sequence of interest [384]. cDNA microarrays are less expensive than oligonucleotide arrays, which along with the greater specificity they offer over oligonucleotide arrays makes cDNA microarrays more popular in biological research. However, cross-hybridisation of homologous sequences is problematic when using cDNA microarrays. It is also difficult, if not impossible, for researchers to have any control over the design of probes and to manage the large number of cDNA libraries [384].

In oligonucleotide arrays, the oligonucleotides can be synthesised either in plates or directly on a glass surface. The probes are designed to target a unique

gene sequence so that cross-hybridisation is minimised to a degree that depends on the available sequence information. Their coverage, consistency, better quality control and uniformity of probe length are some of the advantages of oligonucleotide arrays, even though they are more expensive [180].

There are two ways to detect the signal intensity of probes on an array following hybridisation. In cDNA arrays, a two-colour fluorescence hybridisation scheme is used in which control and target RNA samples are each labelled with one of two fluorescent dyes (Cy3 and Cy5) and hybridised on the same array. The target RNA hybridises with the corresponding complementary probes that have been spotted on the array surface. After hybridisation, the relative balance of green and red fluorescence indicates the relative expression level of the control and target RNAs. Hence, gene expression values are reported as the ratio of the two fluorescent values.

A one-colour hybridisation system is used for oligonucleotide arrays, in which target RNAs are labelled with a single colour fluorescent dye and hybridised onto the array. The fluorescent intensity of each spot provides the absolute abundance of the corresponding target RNA [180].

A scanner is used to detect and record the intensity of fluorescence for each spot/area on the microarray slide. The signal intensity associated with each spot (or probe) indicates the abundance of the corresponding sequence in the hybridised sample. The data generated for control and target samples can be used to calculate the ratio of gene expression in target in relation to control samples [257].

If a gene is over expressed in a particular sample, a large amount of that specific mRNA will hybridise to the probe on the microarray slide, generating a bright area. In the same way, an absence of fluorescence shows that the gene is inactive in that particular sample [362]. Most oligonucleotide microarray experiments use different amounts of control and target samples and analyse them on the basis of absolute expression levels rather than ratios [257].

Microarray data

Probe-level data are completely dependent on the type of microarray chip. Microarray chip manufacturers such as Affymetrix [213], Illumina [17], etc. provide vendor-specific pre-processing and normalisation methods to convert the individual probe-level data to a gene expression matrix. These gene expression matrices consist of probe identity as rows, and sample identity as columns. A simple example of a gene expression matrix is given in Table 2.1.

Table 2.1: Example of a gene expression matrix.

Probe-id	Samp1	Samp2	Samp3	Samp4
1053_at	2.29E-06	1.89E-06	3.19E-06	2.92E-06
117_at	1.34E-06	6.28E-06	9.44E-07	2.48E-06
121_at	8.31E-06	2.54E-05	1.59E-05	1.50E-05
1255_g_at	6.05E-06	8.06E-06	2.19E-06	7.11E-06
AFFX-BioB-5_at	4.81E-06	1.38E-05	9.86E-06	8.59E-06
AFFX-BioB-M_at	2.36E-05	5.35E-05	6.63E-05	4.62E-05
AFFX-BioB-3_at	3.38E-07	6.24E-07	1.36E-06	3.49E-07

Probe-id denotes probe identity. **Samp1**, **Samp2**, **Samp3**, **Samp4** are sample identities. The expression values for each probe corresponding to each sample are given in the rows.

Quality assessment

Quality assessment is an important part of microarray analysis. The publicly available data should be quality controlled to avoid problematic probe-level data [65]. The assessment of microarray data is usually performed by comparing suitable numerical summaries across microarrays, which helps to identify poor - or variable - quality sets of arrays.

As tens or hundreds of thousands of measurements can be obtained from a single array, finding a suitable numerical summary is quite challenging. To this end, several methods have been developed based on probe-level and probe set-level information to identify and remove poor-quality probes from further analysis [35, 61, 173, 299, 378, 369].

Data storage

Online databases are available for storing microarray probe-level data along with biological information relating to the data. The National Center for Biological Information (NCBI; http://www.ncbi.nlm.nih.gov) and the European Bioinformatics Institute (EBI; http://www.ebi.ac.uk) are the two largest databases, and they provide not only microarray data but also nucleotide sequence data, protein information, gene descriptions and disease annotations. The most commonly used repositories for microarray gene expression data are NCBI's Gene Expression Omnibus (GEO) [14] and EBI's ArrayExpress Archive [184]. Both of these repositories support the retrieval of microarray raw and processed data for analysis from a variety of organisms.

2.1.2 Meta-analysis of microarray studies

Many studies use microarray data to detect DEGs between different conditions, but these data are highly variable and may result in non-reproducible outcomes [290]. It is therefore recommended to integrate microarray studies that address the same research question. The basis for the integration of these independent small studies is that they may share common information that can be explored with the application of meta-analysis [1].

Hu, Greenwood and Beyene [146] have implemented a meta-analysis approach for gene expression microarray data that calculates the effect size of genes. The effect size of each gene in individual studies is calculated using fold change (FC, a measure of the quantity of change from an initial to a final expression value), t-test, Z-score and p-value, and these results are integrated into a single measure. This method was applied to a lung cancer dataset, from which it identified a set of DEGs. However, this method is applicable only for Affymetrix platforms, not for a combination of different platforms.

A random effects meta-analysis was performed by Ghosh et al. [100] and identified a list of DEGs in prostate cancer in comparison with benign prostate tissue across multiple studies. Kuo et al. [191] conducted a meta-analysis of gene expression data produced from two different microarray platforms and suggested that directly combining measures from different platforms can result in a very low correlation.

A q-value-based statistical model for performing meta-analysis was demonstrated by Rhodes et al. [295] and applied to four prostate cancer datasets. In this method, the q-value is calculated from the ratio between the expected and actual number of occurrences of a gene. A set of common genes was identified that are consistently and significantly dysregulated and have a specific role in relation to prostate cancer.

Fishel, Kaufman and Ruppin [80] developed a rank-based meta-analysis model for the classification of genes. This method ranks genes according to their repeatability frequency: that is, the frequency of appearance of each gene in the different predictive gene sets. The method was applied to two lung cancer datasets and identified a set of genes that play a key role in tumorigenesis in the lung. The robustness of the resulted genes was demonstrated using a third lung cancer dataset that led to the successful classification of tumour and control samples using very few top-ranked genes.

A method called Similarities of Ordered Gene Lists was developed by Yang and Sun [395] and used to identify consistent changes in genes in different types of cancers. This resulted in a list of genes involved in different types of cancers,

most of which are involved in the breakdown of EMC (extra machrochaetae) protein, which regulates angiogenesis, and may be useful as prognostic markers and molecular targets for gene therapy in cancer.

A comparative meta-profiling method was developed by Rhodes et al. [294] to identify gene expression signatures in cancer. The application of the proposed method resulted with a subset of genes from a diverse collection of cancer datasets. A transcriptional profile was characterised from this list of genes and validated on independent datasets.

As meta-analysis is used in a wide range of research fields, this discussion has focussed only on studies that used microarray data and that are relevant to the current study. They all suggest that a well-designed meta-analysis can provide more reliable results than the individual studies.

Although these methods are capable of selecting features from an integrated dataset based on heterogeneous platforms, they cannot deal with genes that are not represented in all datasets. That is, these methods miss those genes that are not present in all the selected datasets. However, *RankProd* [140] and *GeneMeta* [41], which are by far the most popular among the meta-analysis methods, have been chosen from among such methods as a comparison benchmark. They are explained in the following sections in detail.

RankProd

RankProd is a non-parametric meta-analysis tool introduced by Hong et al. [140] to detect DEGs. It is a modified and extended version of the rank product method proposed by Breitling et al. [30] and is arguably the most widely used gene expression meta-analysis tool used for gene expression data. RankProd detects items that are consistently highly ranked in a series of lists. For example, in the case of microarray data meta-analysis, RankProd helps to identify genes that are consistently found in the lists of the most up- or down-regulated genes in selected studies.

FC is used as a scoring criterion to rank and compare both up-and down-regulated genes within each dataset. An overall ranked gene list is produced by aggregating the individual ranks across datasets. This method helps to overcome the heterogeneity among multi-platform datasets, which simplifies the extraction, comparison and integration. The method can integrate data produced using different platforms such as Affymetrix oligonucleotide arrays, two-colour cDNA arrays and other custom-made arrays. RankProd is implemented in R [349] (http://www.r-project.org) as an open source Bioconductor package

[97] (http://bioconductor.org/)and it accepts the pre-processed microarray dataset in a matrix form. The package provides the functions to perform meta-analysis as well as individual data analysis.

Consider a simple, two-channel microarray experiment comparing mRNA levels under two conditions A and B that examines n genes (or other features) in m samples. For each gene s in the i_{th} sample, each examining n_i genes, its combined probability of appearance can be calculated as a rank product. That is, the position of gene s in the list of up-regulated genes in the i_{th} sample is given as:

$$RP_s^{up} = \prod_{i=1}^{m} (r_{si}^{up}/n_i)$$

in which r_{si}^{up} is the rank or position of gene s in the list of up-regulated genes in the i_{th} sample sorted by decreasing FC. That is, $r^{up} = 1$ for the most strongly up-regulated gene, and so on.

In the same way, RP_s^{down} is calculated for the list of down-regulated genes that are sorted by increasing FC: $r^{down} = 1$ for the most strongly down-regulated gene.

When $n_i = n$ for all samples, the rank product can be calculated as the geometric mean:

$$RP_s = (\prod_{i}^{m} r_{si})^{1/m}$$

Genes with the smallest RP values are the most interesting genes and can be selected for further study.

The pairwise fold change (pFC) for both up- and down-regulated genes in each dataset is computed as:

$$T_1^s/C_1^s, T_1^s/C_2^s, ..., T_2^s/C_1^s, ..., T_{n_{T_m}}^s/C_{n_{C_m}}^s$$

in which T_i^s and C_i^s are the expression values of gene s for sample i belonging to either of the experimental conditions T, 'tumour' and C, 'control'; n_{T_m} and n_{C_m} are the number of pFC values per gene. The corresponding pFC ratios are then ranked and are denoted as r_{si}, where $s = \{1, ..., n\}$ genes and $i = \{1, ..., m\}$.

The expression value for each gene in each dataset is independently permuted T times and produces RP_s^t, where $t = \{1, ..., T\}$, through repetition of all the steps given above. A reference distribution is obtained from all RP_s^t

and the adjusted p-value and false discovery rate (FDR) for each gene are calculated.

For the single-channel arrays, RP values are calculated using all possible pairwise comparisons. Hence a significance analysis has to be performed for each gene: the significance level of gene s in sample A and sample B, each examining n genes can be calculated as:

$$RP_s = (r_s^A/n) \times (r_s^B/n)$$

In simple cases, when the number of genes is small, this probability can be directly calculated from the RP values. That is, for m samples and n genes, the probability can be calculated by multiplying each RP value by a factor F, where F is the number of possible products of the ranks of the gene s. For example, consider an experiment with two samples ($m = 2$) and three genes ($n = 3$). If gene s is has rank 2 in sample A and rank 1 in sample B, its RP value will be $RP = (2/3) \times (1/3) = 2/9$. Note that the same gene has the same RP value if it has rank 1 in sample A and rank 2 in sample B. The combined probability is $p = 2RP = 4/9$ when the RP value is multiplied by factor F, which in this example is 2 because there are two possible products (1×2 and 2×1).

RankProd provides a list of up- and down-regulated genes, along with their ranks. The position of a gene in the list indicates the significance of that gene in the samples used in the experiment, which can be further analysed (please refer [30] for more details about the method and its implementation).

GeneMeta

GeneMeta is another popular meta-analysis method designed for same-platform situations. It was introduced by Lusa, Gentleman and Ruschhaupt [221] as an R package based on the fixed and random effects-based meta-analysis method proposed by Choi et al. [41]. The GeneMeta package allows selection of the effect size model (fixed or random) to be used in an analysis, which determines which analysis will be applied. If a fixed effect model does not hold, then a random effects model will need to be fitted. The estimates of the overall effect, μ, will depend on which model is used. Choi et al. [41] suggest using FDR for both fixed and random models, which can be calculated as:

$$FDR = \frac{\frac{1}{B}\sum_b \sum_s p(|\mu_s^{*b}| \geq \mu_i)}{\sum_s p(|z_s| \geq \mu_i)}$$

where B is the number of column-wise permutations performed in each dataset, each represented as $b = 1, 2, ..., B$, and μ_s^* is the average effect size for gene (or other feature) s. The total number of datasets is denoted p, and $p(\cdot)$ is the indicator function (equal to 1 if the condition in parenthesis is true and 0 otherwise). The denominator represents the number of genes that are significant in the data, and the numerator is the expected number of falsely significant genes.

2.2 Conclusion

This chapter has explained microarray technology in detail, as microarray gene expression data are used in the current study to demonstrate the application of our method. An introduction to different types of meta-analysis approaches was also provided. The two most popular, RankProd and GeneMeta, which will be used as a comparison benchmark in Chapter 4 and Chapter 5 were explained in detail.

Chapter 3

Proposed Methods for Feature Selection

In this chapter, the most important recent studies on feature selection related to this work are briefly discussed in Section 3.1. The (α, β)-k Feature Set problem approach is explained in detail in Section 3.2, since the method proposed for feature selection is a variant of this approach. The proposed methods are explained in detail in Section 3.3 and Section 3.4. The mathematical formulations involved in the method are presented in Section 3.5. Finally, a conclusion for the chapter is provided in Section 3.6.

3.1 Feature selection approaches

Feature selection, also known as subset selection, attribute selection or variable selection, is the process of identifying a set of features that are sufficient to distinguish class labels (for instance tumour and normal) in data. Input data such as microarray data, which are used as an application source for the current study, contain a very large number of features; however many features do not provide class separation or display varying performance in achieving this goal. Moreover, in the analysis of microarray datasets, feature selection helps to provide a deeper understanding of the molecular basis of a disease by removing redundant and irrelevant features from the data, which increases the accuracy of classification of new instances. The process of feature selection also helps biologists to investigate a set of genes or features that are closely related with the classes in the considered datasets.

Feature selection approaches can be organised into two categories – supervised and unsupervised [121] – which are explained in the following sections.

3.1.1 Unsupervised methods

An unsupervised approach does not require class information for the samples and can be used to extract information when the biological datasets are incomplete, or to find subgroups of interest. In most cases, this approach is used only when the dataset has missing information or is incomplete. For example, clustering is the most commonly used unsupervised approach when class labels are not known [99]. Principal component analysis and artificial neural networks are two other types of unsupervised learning approaches [66].

The simplest unsupervised approach for the evaluation of features is the use of variance. As gene expression varies under different conditions, a larger variance in expression indicates greater gene relevance. If the variance associated with a gene is very low, then the gene is removed from further analysis. Such selected features have poor discriminatory power for the classification of data, thus variance is mostly used as a component of a broader approach [54].

In 2006, Varshavsky et al. [360] proposed an unsupervised approach based on singular value decomposition (SVD) entropy, which involves factorisation of a real or complex matrix. The contribution of each feature is calculated via the leave-one-out approach and those that contribute more to the data are selected. However, the calculation of SVD entropy for a large number of features is very computationally expensive.

An extended version of ranking-based feature selection using the Laplacian score was proposed by He, Cai and Niyogi [130] in 2005. This approach depends on the observation that two data points probably share information if those points are close to each other.

Few studies have used unsupervised approaches for feature selection; they are more commonly used for exploratory data analysis because they usually provide unreliable features.

3.1.2 Supervised Methods

Supervised methods consider both the data variance and distribution according to sample classes in the data. These methods are further classified as wrapper, embedded and filter methods.

A wrapper method is a supervised approach to performing multivariate gene selection by including the classifier's bias into a search that attempts to construct the most accurate classifiers. The number of possible subsets depends on the number of features in the data, which makes the search infeasible when there is a large number of features. Therefore, wrapper methods are best preceded by a further step to avoid the necessity for an exhaustive search of all

possible subsets. Wrapper methods depend on the learning algorithm used and require extensive computation for the search of accurate features [163].

Due to these problems, most wrapper methods that are used to analyse microarray data employ heuristic searches, among which evolutionary algorithms are the most popular. In a study of the molecular classification of cancer, Duval and Hao [64] used a memetic algorithm [249, 250, 251, 262] that employs information from a support vector machine (SVM). That is, they used SVM ranking-based fitness and cross-over operations to select features. The resulting features were refined using an iterated local search procedure [218]. Li et al. [204], Ooi and Tan [269] and Jirapech-Umpai and Aitken [158] used a genetic algorithm that employs k nearest neighbours to select and evaluate features.

Embedded methods differ from other feature selection methods in that they include learning interaction with the classification approach, and are less computationally intensive than wrapper methods. In the case of embedded methods, the learning part cannot be separated from the feature selection part, highlighting that the structure of the class of functions under consideration plays an important role. However, in the case of microarray studies, embedded methods have high computational complexity due to the large number of features in microarray data.

An embedded approach such as a 'random forest' has the ability to assess the relevance of a single gene based on a set of discriminative features [156, 52].

An SVM-based feature selection method known as support vector machine–recursive feature elimination (SVM-RFE) that was proposed by Guyon et al. [121] trains the classifier by optimising the weights of the features, and then ranking all features and eliminating those with the lowest ranks. A generalisation of Monte Carlo and tree classifier approaches was developed by Draminski et al. [59]. The method selects the features that take part in the classification process more often.

A *Filter method* is a supervised approach used to analyse the intrinsic properties of data by ignoring the classifier. Most filter approaches are based on two operations: ranking and subset selection. In ranking, the importance of each feature is evaluated without considering the potential interactions among the elements of the joint set. In subset selection, the final subset of features to be selected is provided. Filter-based approaches are the most commonly used approaches for microarray data analysis. This approach helps to evaluate DEGs or other features and rank them according to their ability to distinguish the classes. DEGs show a certain level of expression under one set of conditions

and a significantly different distribution under other conditions. Filter methods can be subdivided into univariate and multivariate methods.

Univariate methods rank features by assigning a score value to each. Ben-Dor et al. [5] proposed a method called TNoM (threshold number of misclassification) that uses the FC differences between classes for features in the data to set a threshold and rank the features. Two parametric variations of TNoM were proposed by Jafari and Azuaje [155] in 2006. Also, Baldi and Long [13] and Fox and Dimmic [85] proposed Bayesian approaches based on t-tests to overcome deficiencies related to the low replication, in a statistically consistent way.

More general approaches like the use of Wilcoxon rank sum as proposed by Thomas et al. [351], BSS/WSS [63] and RankProd [30] can be used when no distribution assumptions can be made.

Multivariate methods use groups of features instead of individual features, and define a multivariate fitness function to identify the combination of features to evaluate. The simplest multivariate approach [26] uses only bivariate interactions whereas others use more complex and high-order interactions [124, 370, 396, 98, 226, 387]. More complex approaches such as minimum-redundancy maximum-relevance [55] and uncorrelated shrunken centroid [397] have also been successfully exploited. Finally, the fully multivariate (α, β)-k Feature Set problem approach has the ability to control the robustness of the solution and the number of features to be selected and has been used successfully in various studies, as discussed next.

3.2 The $(\alpha, \beta) - k$ Feature Set Problem approach

The combinatorial optimisation-based (α, β)-k Feature Set problem approach proposed by Cotta, Sloper and Moscato [46] is a generalisation of the k-Feature Set problem. It is a supervised feature selection approach to select a significant set of features that can collectively maximise inter-class discrimination and intra-class coherency [241]. The method seeks to differentiate all sample pairs that belong to different classes by selecting a minimum set of genes that does not necessarily present a uniform expression level across samples in each class but collectively provides the maximum amount of evidence.

The k-feature set problem has been shown to be NP-hard [46, 49] and W[2]-hard [45]. Thus the (α, β)-k Feature Set problem is also NP-hard and W[2]-hard.

3.2.1 The problem definition

The decision version of the (α, β)-k Feature Set problem is presented below, where $\mathbb{B}$ represents the set of binary values, that is, $b \in \{0, 1\}$; n is the number of features, m is the number of samples and the tuple y is the class label.

(α,β)-k-FEATURE SET:

Instance: A set $X = \{x_i \mid x_i \in \mathbb{B}^n \wedge 1 \leq i \leq m\}$, a tuple $y \in \mathbb{B}^m$, integers $\alpha > 0$, $\beta \geq 0$, $k > 0$.

Parameter: $\alpha + \beta + k$.

Question: Is there a set $I \subseteq \{1, \ldots, n\}$ with $|I| \leq k$ such that for all $i, j \in \{1, \ldots, m\}$

- if $y_i \neq y_j$ there exists $I^{\alpha}_{i,j} \subseteq I$ with $|I^{\alpha}_{i,j}| \geq \alpha$ such that $x_{i,s} \neq x_{j,s}$ for all $s \in I^{\alpha}_{i,j}$,
- if $y_i = y_j$ there exists $I^{\beta}_{i,j} \subseteq I$ with $|I^{\beta}_{i,j}| \geq \beta$ such that $x_{i,s} = x_{j,s}$ for all $s \in I^{\beta}_{i,j}$?

In other words, the (α, β)-k Feature Set problem aims to determine if there exists a set of k features that explains the divergence within the samples by maximising the similarities between samples that belong to the same class and the differences between two samples belonging to different classes. Consider a discrete binary valued matrix M with m samples and n features, and an array of sample class labels L. The matrix M has the values $x_{j,s}$ for each sample j and feature s. The array L is of size m with class l_j for each sample j. The problem is defined by three parameters α, β and k, where α represents the minimum number of features that explains the difference between any pair of samples from different classes, β represents the minimum number of features that explains the similarities between any pair of samples from the same class, and k is the number features to be selected.

Consider the example given in Table 3.1 for a better understanding of the problem.

Table 3.1: An example of a numerical data with seven features and five samples.

Features	**Samp1**	**Samp2**	**Samp3**	**Samp4**	**Samp5**
Gene A	0	0	1	0	0
Gene B	1	1	1	0	1
Gene C	1	1	0	1	0
Gene D	1	1	1	0	0
Gene E	0	1	0	0	0
Gene F	0	1	1	1	0
Gene G	0	1	1	0	0
Class	N	N	D	D	D

Features is the features (probes or genes) present in the data. **Samp1, Samp2, Samp3, Samp4, Samp5** are the samples present in the data that belongs to the class normal (N) or disease (D).

Each row in Table 3.1 corresponds to a feature or a gene and each column corresponds to a sample. The values in the rows represent the level of gene expression for each gene for each sample. The (α, β)-k Feature Set problem is defined in terms of Boolean variables so some discretisation has been applied before and samples belong to one of two classes, which is given in the last row of Table 3.1.

In order to describe the problem, all possible pairs of samples are constructed and transformed the gene expression level to a coverage to indicate whether the gene is capable to explain that particular pair of samples, shown in Table 3.2.

In Table 3.2, the first column gives all possible pairs of samples and the columns 2 to 6 represent the features, in this case it is gene.The entries, 'T' and 'F', are the notation to show whether a particular gene can distinguish the corresponding pair of samples. These entries are the response to the question of whether the gene expression level of a feature is significantly different for the two samples that belong to different classes or the value of that feature is the same for two samples that belong to the same class. The entry 'T' in Table 3.2 shows that the particular feature can discriminate the corresponding sample pairs and 'F' shows that it cannot. For example, in Table 3.2 the entry of Gene A for the sample pair (1, 4) is 'F' indicating that Gene A cannot discriminate the samples 1 and 4, since the expression values for Gene A is not significantly different for samples 1 and 4 in Table 3.1. At the same time Gene A can distinguish the samples 1 and 3 as the expression values for Gene A are significantly different for samples 1 and 3 in Table 3.1. At the same time, the

Table 3.2: Transformed data into a coverage

Sample Pairs	Gene A	Gene B	Gene C	Gene D	Gene E	Gene F	Gene G
(1,3)	T	F	T	F	F	T	T
(1,4)	F	T	F	T	F	T	F
(1,5)	F	F	T	T	F	F	F
(2,3)	T	F	T	F	T	F	F
(2,4)	F	T	F	T	T	F	T
(2,5)	F	F	T	T	T	T	T
(1,2)	T	T	T	T	F	F	F
(3,4)	F	F	F	F	T	T	F
(3,5)	F	T	T	F	T	F	F
(4,5)	T	F	F	T	T	F	T

Sample Pairs are the possible pairs of samples, the upper part is the pairs of samples that belong to different classes and lower part is the pair of samples that belong to same class. **Gene A, B, C, D, E, F, G** are the different features present in the dataset and the entries 'F' and 'T' indicate whether the gene distinguishes the corresponding pair of samples.

entry of Gene A is 'T' for sample pair (1,2) that belongs to the same class is 1, showing that the gene expression level for Gene A in sample 1 and 2 is same. In such a way Table 3.2 is arranged by comparing the gene expression values of each pair of samples. The aim here is to find a k feature set that can explain sample pairs that belong to different classes by at least α features and sample pairs that belong to the same class can be explained by at least β features of the k feature set.

In our application of the methodology, datasets are discretised by the application of Fayyad–Irani's supervised, multi-interval, class-entropy discretisation prescription (refer 3.5 for more details). This procedure has the additional benefit of providing a clear prescription (the Minimum Description Length principle) to decide whether a feature is informative or not for the classes considered, and hence it helps in reducing the problem instance size. But it is not exclusive of other approaches; for example, discretisation of values can be performed in a simplistic unsupervised way by mean value discretisation, or by binning in a small number of ordinal bins. Furthermore, discretisation itself may be avoided altogether, and values compared based on the observed data variance.

3.2.2 The graph representation

The representation of the problem as a graph makes it easy to formulate the problem mathematically. Since it is possible to reduce the basic k-Feature Set problem to the Red-Blue Dominating Set problem on bipartite graphs [46],

(α, β)-k -Feature Set problem can also be represented as a graph optimisation problem.

We can build a bipartite graph $G(V, E)$ from Table 3.2 with three different nodes, shown in Figure 3.1. Each yellow node represents feature (gene) node v_s for each feature s, each blue node represents the α node $v_{i,j}$ for each sample pair (i, j) that belong to different classes (in this case, normal and disease) and each white node represents the β node $v_{i,j}$ for each sample pairs (i, j) that belong to same class.

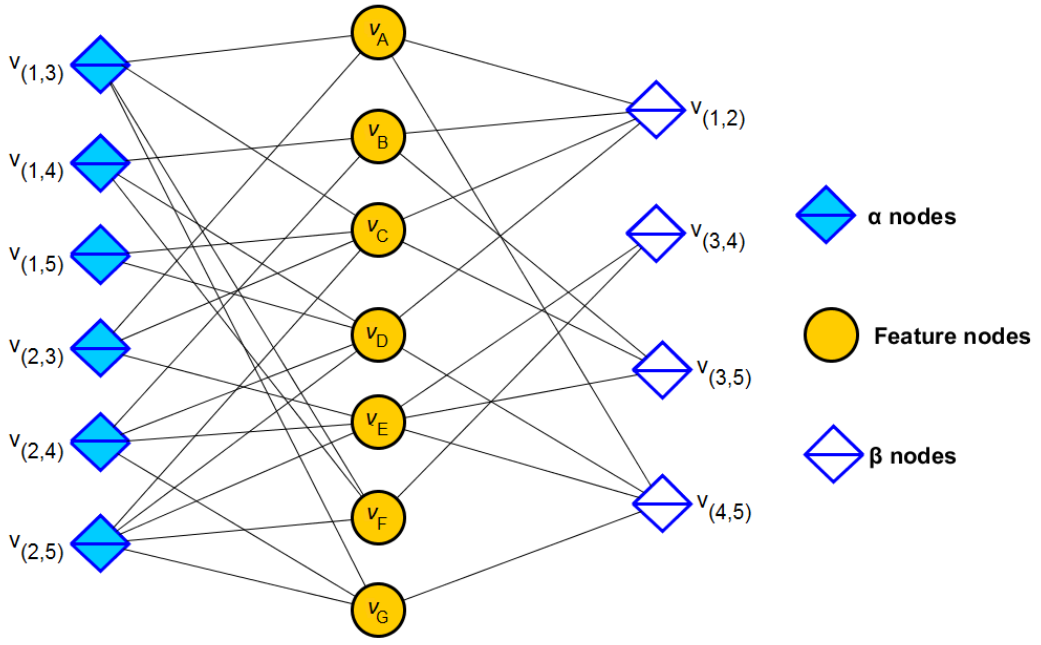
The set of feature nodes are denoted as $\mathcal{F}$, the set of α nodes as $\mathcal{A}$ and the set of β nodes as $\mathcal{B}$. An α edge e_{sij} exists from a feature node v_s (yellow node) to the α node $v_{i,j}$ (blue node) when the gene expression values of that gene is different for that particular sample pair. That is,

$$e_{sij}; e(v_s, v_{i,j}); \quad v_s \in \mathcal{F}, \; v_{i,j} \in \mathcal{A}, \quad x_{i,s} \neq x_{j,s}$$

Similarly, there exists a β edge e_{sij} from a feature node v_s to a β node $v_{i,j}$ if the expression values for that feature do not differ for that particular pair of samples,

$$e_{sij}; e(v_s, v_{i,j}); \quad v_s \in \mathcal{F}, \; v_{i,j} \in \mathcal{B}, \quad x_{i,s} = x_{j,s}$$

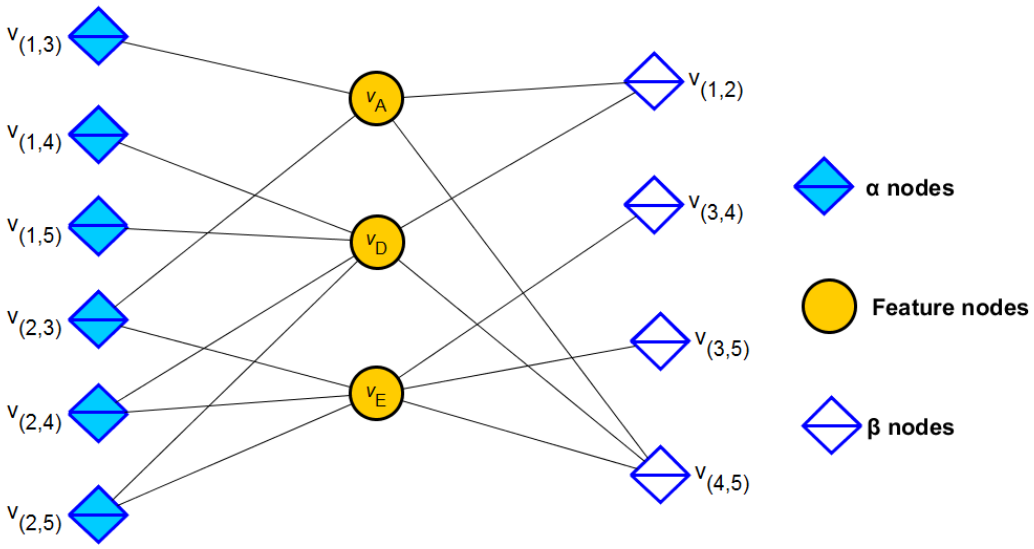
Figure 3.1: Graph representation of the (α, β)-k Feature Set problem instance created from Table 3.2. Blue nodes represent sample pairs that belongs to different classes (α node), yellow nodes represent the feature node and white nodes represent sample pairs that belong to same class (β node). The figure adapted from [20]



For instance, node v_A (Gene A) has an edge to node $v_{(1,3)}$ (sample pair (1,3)) indicating that Gene A has significantly different expression in samples 1 and 3. Also, an edge exists between node v_A (Gene A) and β node $v_{(1,2)}$ (sample pair (1,2)) showing that Gene A has the same expression level in samples 1 and 2. The aim here is to search for a set of k features that can 'explain' the sample pairs that belong to different classes by at least α features and the sample pairs that belong to the same class can be explained by at least β features of the k feature set. For example, a feasible solution for the instance given in Figure 3.1 for $\alpha = 1$ and $\beta = 1$ can be $\{A, D, E\}$, shown in Figure 3.2. Note that $\{C, D, E\}$, $\{B, C, E\}$, $\{C, D, F\}$, $\{D, E, F\}$, $\{C, E, F\}$ are the other possible solutions with three features.

Figure 3.2: An optimal solution for the (α, β)-k Feature Set problem instance from Figure 3.1 with $\alpha = 1$ and $\beta = 1$. The figure was adapted from [20]

.



3.2.3 Previous studies using the (α, β)-k Feature Set problem approach

The (α, β)-k Feature Set problem approach has been used to solve a range of practical problems in reasonable time. In 2004, Cotta, Sloper and Moscato [46] applied the approach to a microarray dataset containing different types of diffuse large B-cell lymphoma samples to identify a subset of genes that can discriminate the different classes. In 2005, the (α, β)-k Feature Set problem

approach was applied to a historical dataset collected during an earlier United States presidential election to predict its outcome [252].

Berretta, Mendes and Mascato [21] used the (α, β)-k Feature Set problem approach to select genes that enable molecular classification of cancer samples. Mendes, Scott and Moscato [233] used this approach to identify molecular portraits of prostate cancer with different Gleason scores. Biomarkers for prostate cancer and AD have been identified via this approach respectively by Gomez Ravetti, Berretta and Moscato [103] and Gomez Ravetti and Moscato [104].

Berretta, Costa and Moscato [20] used the approach to search for high-quality solutions to sequential ordering of expression profiles. A study on multiple sclerosis was performed using this approach by Riveros et al. [300].

The advantage of the (α, β)-k Feature Set problem approach over other feature selection methods is that it seeks to differentiate all sample pairs from different classes by selecting a minimum set of genes that do not necessarily present a uniform expression level across samples in each class but collectively provide the maximum amount of evidence. In contrast, rank methods that score and order genes according to their differential expression across classes bring together gene sets that may not work as a signature, particularly in complex diseases whose molecular characterisation may present subgroups.

The main disadvantage of the (α, β)-k Feature Set problem approach is the complexity of the computational problems involved. As the problems are NP-complete, it is very unlikely that a method exists to solve the problem more efficiently.

The (α, β)-k Feature Set problem approach has been defined having first in mind a single uniform dataset, and conceived in this ways, it is not readily applicable to the case of integrated datasets. An extended version of the approach is required that can handle an integrated dataset in a consistent manner and selects features that differentiate sample pairs across datasets. The application of an (α, β)-k Feature Set problem-based approach to meta-analysis will help to identify the best set of features from a combined dataset, allowing researchers to establish which genetic pathways are involved in the development of a disease. This thesis addresses the need for such an (α, β)-k Feature Set problem-based approach for meta-analysis.

3.3 The Coloured (α, β)-k Feature Set problem approach

'Coloured' here stems from the assumption that samples belong to different categories (in this case different platforms/cohorts/datasets/sample groups) and are given a colour unique to the particular category in which they belong. The approach restricts the original (α, β)-k Feature Set problem definition by now only considering discrimination between samples with same colour. For instance, in the simplest case, each dataset (as a category) is given a unique colour and the method restricts the problem in creating sample pairs from datasets labelled with a different colour.

3.3.1 The problem definition

The decision version of the Coloured (α, β)-k -Feature Set problem is presented below. In what follows, let $\mathbb{B}$ represent the set of binary values, i.e. $b \in \{0, 1\}$; let n be the number of features and m the number of samples, the tuple y be the class labels of the samples and p be the number of datasets. $I^{\alpha}_{i,j}$ is the subset of features that cover the α node and $I^{\beta}_{i,j}$ is the subset of features that cover the β nodes. $x_{i,s}$ is the expression value of feature s for each sample i. Let c be the *colour* of the dataset/sample group that are uniquely assigned, $c \in \{1, ..., p\}$ and $c(j)$ is the colour of that dataset in which the sample j belongs to.

COLOURED (α,β)-k-FEATURE SET:

Instance: A set $X = \{x_i \mid x_i \in \mathbb{B}^n \wedge 1 \leq i \leq m\}$, a colouring function $c : \{1, \ldots, m\} \to \{1, \ldots, p\}$, a tuple $y \in \mathbb{B}^m$, integers $\alpha > 0$, $\beta \geq 0$, $k > 0$.

Parameter: $\alpha + \beta + k$.

Question: Is there a set $I \subseteq \{1, \ldots, n\}$ with $|I| \leq k$ such that for all $i, j \in \{1, \ldots, m\}$ where $c(i) = c(j)$

- if $y_i \neq y_j$ there exists $I^{\alpha}_{i,j} \subseteq I$ with $|I^{\alpha}_{i,j}| \geq \alpha$ such that $x_{i,s} \neq x_{j,s}$ for all $s \in I^{\alpha}_{i,j}$,
- if $y_i = y_j$ there exists $I^{\beta}_{i,j} \subseteq I$ with $|I^{\beta}_{i,j}| \geq \beta$ such that $x_{i,s} = x_{j,s}$ for all $s \in I^{\beta}_{i,j}$?

In words, the Coloured (α, β)-k -Feature Set problem is defined by three parameters α, β and k, where α represents the minimum number of features

that explain the difference between any pair of samples of the same colour that belong to different classes, β represents the minimum number of features that explains the similarities between any pair of samples of the same colour that belongs to the same class and k is the number of features that explain the divergence within the samples by maximising the similarities between the samples belongs to the same classes of same colour and the differences between two samples belongs to different classes of the same colour.

Consider the examples – using blue and orange datasets – given in Table 3.3, Table 3.4 for a better understanding of the problem.

Table 3.3: An example of numerical data for blue dataset (hence the suffix 'b' in the sample names) with five features and five samples.

Features	**Samp1b**	**Samp2b**	**Samp3b**	**Samp4b**	**Samp5b**
Gene A	0	1	1	1	1
Gene B	1	1	0	1	0
Gene C	1	0	1	1	0
Gene D	0	1	1	0	1
Gene E	1	0	1	0	1
Class	N	N	D	D	D

Features is the features (probes or genes) present in the data. **Samp1b, Samp2b, Samp3b, Samp4b, Samp5b** are the samples present in the blue dataset that belong to normal and disease samples.

Table 3.4: An example of numerical data for an orange dataset (hence the suffix 'o' in the sample names) with five features and five samples.

Features	**Samp1o**	**Samp2o**	**Samp3o**	**Samp4o**	**Samp5o**
Gene A	1	1	0	1	1
Gene B	0	1	1	1	1
Gene C	0	1	0	1	0
Gene D	1	0	1	1	0
Gene E	1	1	0	1	1
Class	N	N	D	D	D

Features is the features (probes or genes) present in the data. **Samp1o, Samp2o, Samp3o, Samp4o, Samp5o** are the samples present in the orange dataset that belong to normal and disease samples.

Each row in Table 3.3 and Table 3.4 corresponds to a feature or a gene and each column corresponds to a sample. The values in the rows are the gene expression levels for each gene for the corresponding samples. The values 0 and 1 is used to represent the gene expression levels to simplify the problem. The last row in each table represents the class of each sample. Note that there is no common samples between blue and orange dataset and is restricted the problem in making sample pairs from different dataset.

To describe the problem, all possible sample pairs are constructed from Table 3.3 and Table 3.4 and expression levels for each feature are compared for all sample pairs. The gene expression levels are then transformed to a coverage to indicate whether a gene is capable to explain the corresponding sample pair. The transformed values to a coverage are given in Table 3.5. The

Table 3.5: Transformed value to a coverage for blue and orange datasets.

Colour	Sample pairs	Gene A	Gene B	Gene C	Gene D	Gene E
b	1b,3b	T	T	F	T	F
	1b,4b	T	F	F	F	T
	1b,5b	T	T	T	T	F
	2b,3b	F	T	T	F	T
	2b,4b	F	F	T	T	F
	2b,5b	F	T	F	F	T
	1b,2b	F	T	F	F	F
	3b,4b	T	F	T	F	F
	3b,5b	T	T	F	T	T
	4b,5b	T	F	F	F	F
o	1o,3o	T	T	F	F	T
	1o,4o	F	T	T	F	F
	1o,5o	F	T	F	T	F
	2o,3o	T	F	T	T	T
	2o,4o	F	F	F	T	F
	2o,5o	F	F	T	F	F
	1o,2o	T	F	F	F	T
	3o,4o	F	T	F	T	F
	3o,5o	F	T	T	F	F
	4o,5o	T	T	F	F	T

Colour represents the dataset, blue-b and orange-o, selected for meta-analysis. **Sample Pairs** are the possible pairs of samples. **Gene A, B, C, D, E, F** the different genes present in the dataset.

first column represents the colour of the datasets, b (blue) and o (orange). The second column gives all possible pairs of samples for both *'b'* and *'o'*.

The entries, 'T' for True and 'F' for False, in Table 3.5 are notations to indicate whether a particular gene can distinguish a particular pair of samples. In other words, 'T' and 'F' in Table 3.5 indicate the response to the question of whether the two samples belong to different classes and have different values or to the same class and have the same value in Table 3.3 and Table 3.4.

For example, in Table 3.5 the 'T' entry for Gene A for sample pair ($1b$, $3b$) in *'b'* indicates that Gene A can discriminate samples $1b$ and $3b$ because the expression values for Gene A are significantly different in these two samples in Table 3.3. In contrast, Gene A cannot distinguish samples $2b$ and $3b$ in *'b'* as the expression values for Gene A are not significantly different for samples $2b$ and $3b$ in Table 3.3. The aim here is to find a k feature set that can explain the sample pairs that belong to different classes and the same colour by at least α features and the sample pairs that belongs to the same class and colour can be explained by at least β features of the k feature set.

3.3.2 The graph representation

A graph $G(V, E)$ can be constructed with three different types of nodes; a feature node v_s for each feature s, $s \in \{1, ..., n\}$, an α node $v_{i,j}$ for each sample pair (i, j) that belongs to the same colour and different classes

As mentioned previously, the set of feature nodes is denoted as $\mathcal{F}$, the set of α nodes as $\mathcal{A}$ and the set of β nodes as $\mathcal{B}$.

There exists an α edge $e_{sij}$from a feature node v_s to an α node $v_{i,j}$ when the expression value of given feature is different for that particular sample pair,

$$e_{sij}; e(v_s, v_{i,j}); \quad v_s \in \mathcal{F}, \ v_{i,j} \in \mathcal{A}, \quad x_{i,s} \neq x_{j,s}$$

Similarly, there exists a β edge e_{sij} from a feature node to a β node if the value of that feature is the same for both the members of that particular sample pairs,

$$e_{sij}; e(v_s, v_{i,j}); \quad v_s \in \mathcal{F}, \ v_{i,j} \in \mathcal{B}, \quad x_{i,s} = x_{j,s}.$$

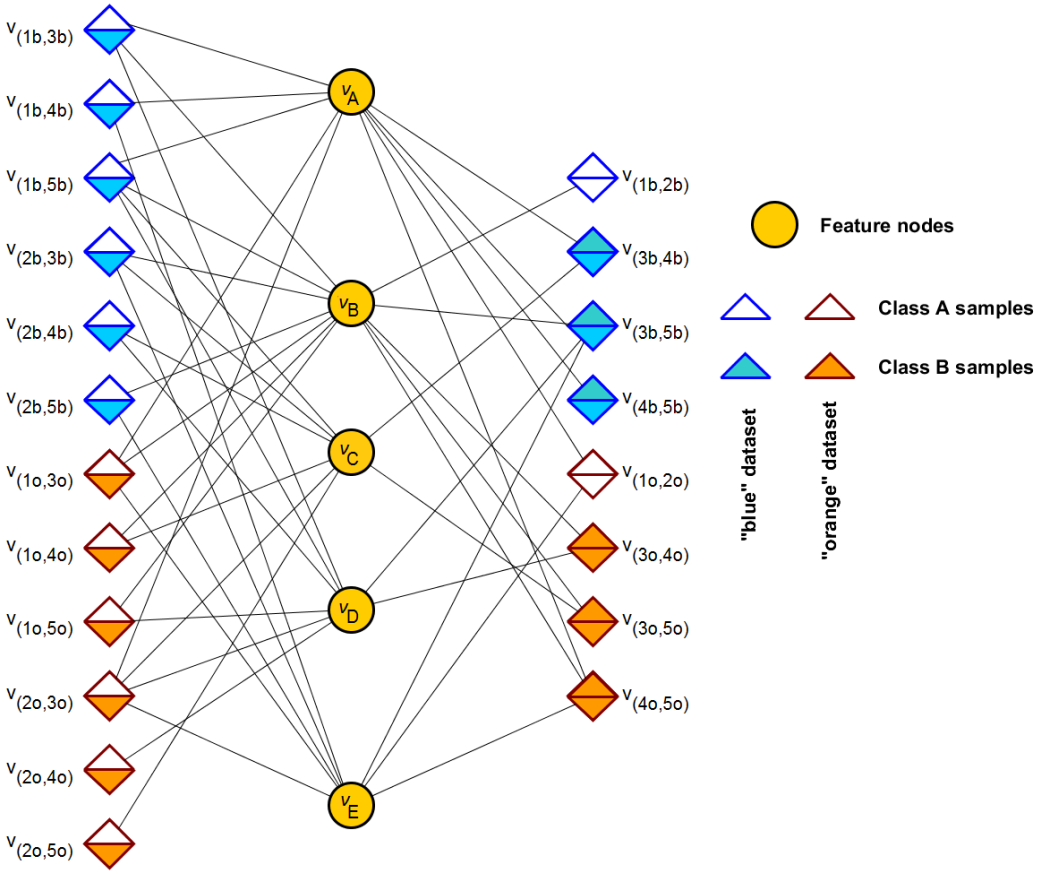
Whenever there exists an edge e_{sij}, indicating that feature s can 'explain' the sample pair (i, j).

Consider the graph given in Figure 3.3 that was created from Table 3.5.

In the graph Figure 3.3, each yellow node represents a feature(gene) node for each feature, each blue node represents the node for each sample pairs that belongs to the 'blue' dataset *'b'* and each orange node represents the node for each sample pairs that belong to the 'orange' dataset *'o'*. An edge is added according to the entries in Table 3.5. For example, node v_A (Gene A) has an edge to α node $v_{(1b,3b)}$ (sample pair (1b,3b)) because Gene A has a significantly different expression value for samples 1b and 3b. Also, an edge exists between node v_A (Gene A) and β node $v_{(3b,4b)}$ (sample pair(3b,4b)) indicating that Gene A has the same expression levels in samples 3b and 4b. In this way a search can be performed for a set of k features that cover all sample pairs. Note that there is no α or β nodes are mixed with blue and orange colour, because the approach restricts the problem to make pairs of samples that belong to different datasets or colours. A feasible solution for the instance provided in Figure 3.3 with $\alpha = 1$ and $\beta = 1$ can be $\{A, B, C, D\}$, shown in Figure 3.4. Note that there may be other possible solutions with four features.

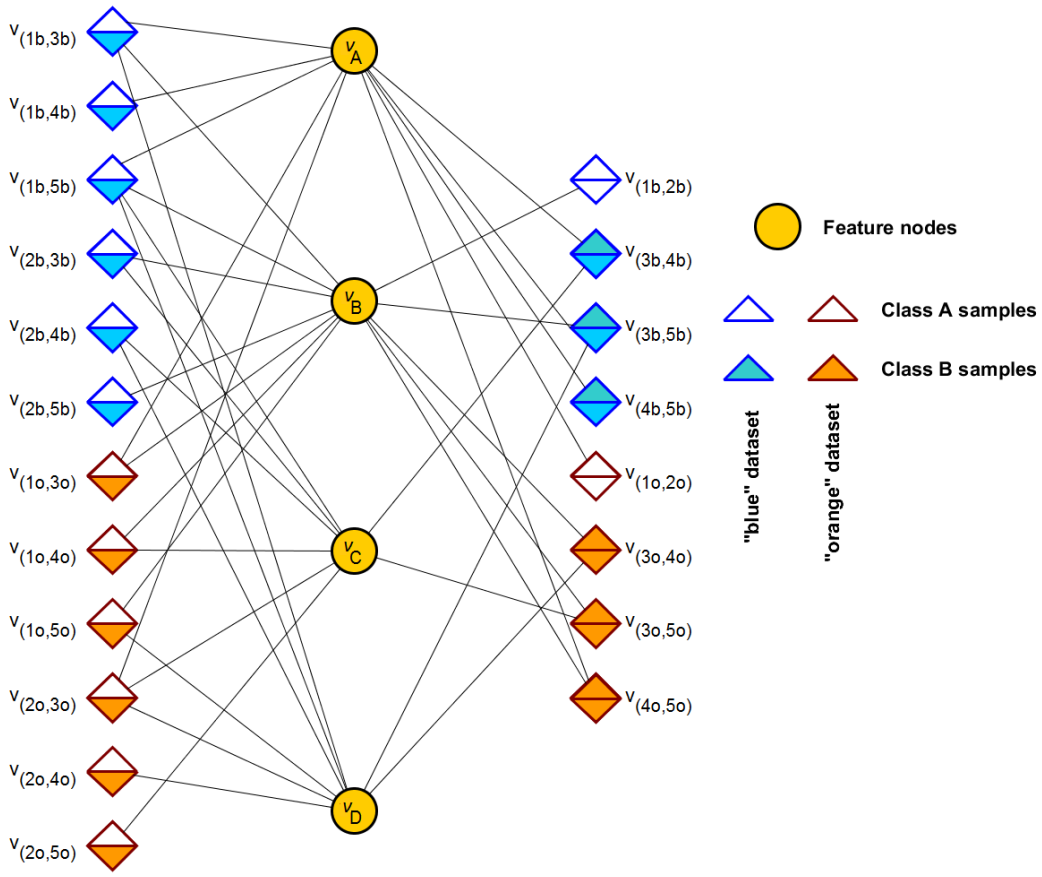
The simplest version of the Coloured (α, β)-k Feature Set problem approach aims to maximise the cover of α and β nodes with the minimum number of k features. The approach presented here is just one of several possible variants in the construction of a graph that gives rise to the optimisation models. It

Figure 3.3: Graph representation of the Coloured (α, β)-k Feature Set problem created from Table 3.5. Each blue nodes represents sample pairs that belong to the 'blue' dataset, *'b'*; each orange node represents sample pairs that belong to the 'orange' dataset, *'o'* and each yellow node represents the features.



might be desirable to restrict the creation of α or β pairs according to other constraints, for example, by only allowing for certain combinations of colours, or by requiring that α nodes be created from samples of the same colour and β nodes from samples of different colours.

Figure 3.4: An optimal solution for the Coloured (α, β)-k Feature Set problem instance from Figure 3.3 with $\alpha = 1$ and $\beta = 1$.



3.4 The Generalised (α, β)-k Feature Set problem approach

As explained in Section 3.3 by considering discrimination between samples with same colour and the problem is restricted to make pairs of samples that belong to different colours. In this case there are no samples shared among datasets. However datasets can sometimes share samples. In which case, shared samples should be allowed to make pairs with samples that belong to other datasets as well. In this case colour can not be used to represent a dataset, which leads to a generalised version of the (α, β)-k Feature Set problem approach.

3.4.1 The problem definition

The decision version of the generalised (α, β)-k -Feature Set problem is presented below. In what follows, let $\mathbb{B}$ represent the set of binary values, i.e. $b \in \{0, 1\}$; let n be the number of features and m the number of samples and the tuple y be the class labels for the samples.

GENERALISED (α,β)-k-FEATURE SET:

Instance: A set $X = \{x_i \mid x_i \in \mathbb{B}^n \wedge 1 \leq i \leq m\}$, a function $f : \{1,\ldots,m\} \times \{1,\ldots,m\} \to \mathbb{B}$, a tuple $y \in \mathbb{B}^m$, integers $\alpha > 0$, $\beta \geq 0$, $k > 0$.

Parameter: $\alpha + \beta + k$.

Question: Is there a set $I \subseteq \{1,\ldots,n\}$ with $|I| \leq k$ such that for all $i,j \in \{1,\ldots,m\}$ where $f(i,j) = 1$

- if $y_i \neq y_j$ there exists $I^{\alpha}_{i,j} \subseteq I$ with $|I^{\alpha}_{i,j}| \geq \alpha$ such that $x_{i,s} \neq x_{j,s}$ for all $s \in I^{\alpha}_{i,j}$,
- if $y_i = y_j$ there exists $I^{\beta}_{i,j} \subseteq I$ with $|I^{\beta}_{i,j}| \geq \beta$ such that $x_{i,s} = x_{j,s}$ for all $s \in I^{\beta}_{i,j}$?

In words, the Generalised (α,β)-k -Feature Set problem is defined by three parameters α, β and k, where α represents the minimum number of features that explain the difference between any pair of samples that belong to different classes, β represents the minimum number of features that explains the similarities between any pair of samples that belongs to the same class and k is the number of features that explain the divergence among the samples by maximising the similarities between samples belonging to the same class and the differences between two samples belonging to different classes.

Consider the examples – using 'first' (hence the suffix 'f' for the samples) and 'second' dataset (hence the suffix 's' for the samples) – given in Table 3.6 and Table 3.7 to better understand the problem.

Table 3.6: An example of numerical data for *first* dataset with five features and five samples.

Features	**Samp**1f	**Samp**2f	*Samp*3f^*	**Samp**4f	**Samp**5f
Gene A	0	1	1	1	1
Gene B	1	1	0	1	0
Gene C	1	0	1	1	0
Gene D	0	1	1	0	1
Gene E	1	0	1	0	1
Class	N	N	D	D	D

Features is the features (probes or genes) present in the data. **Samp**1f**,** **Samp**2f**,** *Samp*3f^***,** **Samp**4f**,** **Samp**5f are the samples in the first dataset that belong to the classes normal and disease.

To describe the problem, all possible sample pairs are constructed from Table 3.6 and Table 3.7 and the expression level of each feature is compared

Table 3.7: An example of a numerical data for *second* dataset with five features and five samples.

Features	**Samp**1*s*	**Samp**2*s*	*Samp*3*s**	**Samp**4*s*	**Samp**5*s*
Gene A	1	1	0	1	1
Gene B	0	1	1	1	1
Gene C	0	1	0	1	0
Gene D	1	0	1	1	0
Gene E	1	1	0	1	1
Class	N	N	D	D	D

Features is the features (probes or genes) present in the data. **Samp**1*s*, **Samp**2*s*, *Samp*3*s**, **Samp**4*s*, **Samp**5*s* are the samples that belongs to the classes normal and disease and is present in the second dataset.

for each sample pair. As mentioned before, the gene expression levels are transformed to a coverage given in Table 3.8. The first column represents the datasets, **f** (first dataset) and **s** (second dataset). The second column gives all possible pairs of samples for both datasets. Note that sample 3, common to both datasets, is allowed to make pairs with all samples in both datasets. For instance, according to the Coloured (α, β)-k Feature Set problem approach, sample 3 from the first dataset is not allowed to make pair with samples in the second dataset. However in the generalised (α, β)-k Feature Set problem approach, sample 3 in the first dataset can make pairs with samples in the second dataset and sample 3 from the second dataset can make pairs with samples in the first dataset.

The entries, 'T' for True and 'F' for False, in Table 3.8 are the notations to show whether a particular gene can distinguish the corresponding pair of samples (refer Subsection 3.3.1 for more details).

Table 3.8: Transformed value to a coverage for datasets f (first) and s (second).

Dataset	Sample pairs	Gene A	Gene B	Gene C	Gene D	Gene E
f	(1f,3f)	T	T	F	T	F
	(1f,4f)	T	F	F	F	T
	(1f,5f)	T	T	T	T	F
	(2f,3f)	F	T	T	F	T
	(2f,4f)	F	F	T	T	F
	(2f,5f)	F	T	F	F	T
	(1f,2f)	F	T	F	F	F
	(3f,4f)	T	F	T	F	F
	(3f,5f)	T	T	F	T	T
	(4f,5f)	T	F	F	F	F
s	(1s,3s)	T	T	F	F	T
	(1s,4s)	F	T	T	F	F
	(1s,5s)	F	T	F	T	F
	(2s,3s)	T	F	T	T	T
	(2s,4s)	F	F	F	T	F
	(2s,5s)	F	F	T	F	F
	(1s,2s)	T	F	F	F	T
	(3s,4s)	F	T	F	T	F
	(3s,5s)	F	T	T	F	F
	(4s,5s)	T	T	F	F	T
f&s	(1s,3f)	F	F	T	F	F
	(2s,3f)	F	T	F	T	F
	(3f,4s)	T	F	T	T	T
	(3f,5s)	T	F	F	F	T
	(1f,3s)	F	F	T	T	T
	(2f,3s)	T	F	F	F	F
	(3s,4f)	F	T	F	F	F
	(3s,5f)	F	F	T	T	T

Dataset represents the datasets that are selected for the analysis,that are f-first dataset and s-second dataset. f&s gives the sample pairs that sample 3 make from both datasets. **Sample Pairs** are the possible pairs of samples. **Gene A, B, C, D, E, F** the genes present in the dataset.

3.4.2 The graph representation

In this case there is no colour assigned, the comparison of expression values for two samples for every feature between datasets is described by an adjacency relationship expressed as a matrix: $D = \delta_{i,j}, \delta_{i,j} \in \{0, 1\}$.

A graph $G(V, E)$ can be constructed with three different nodes: a feature (gene) node v_s for each feature s, $s \in \{1, ..., n\}$, an α node $v_{i,j}$ for each permissible pair of samples according to the adjacency relationship $\delta_{i,j} = 1$:

With all these definitions, addition of edges between nodes proceeds as explained in Subsection 3.3.2. Note that the arbitrary comparison relations between samples of different or same classes is performed by defining the adjacency for each sample.

Consider the graph given in Figure 3.5 created from Table 3.8.

An edge is added according to the entries in Table 3.8 (refer subsection 3.3.2). An optimal solution for the Generalised (α, β)-k Feature Set problem instance in Figure 3.5 is given in Subsection 3.6.

3.5 Mathematical formulation

In this section, the Coloured (α, β)-k Feature Set problem is presented as an Integer Program with binary variables.

$$min(\sum_{s=1}^{n} f_s)$$

where the variable $f_s = 1$, if the feature s is selected to the k feature set and 0 otherwise.

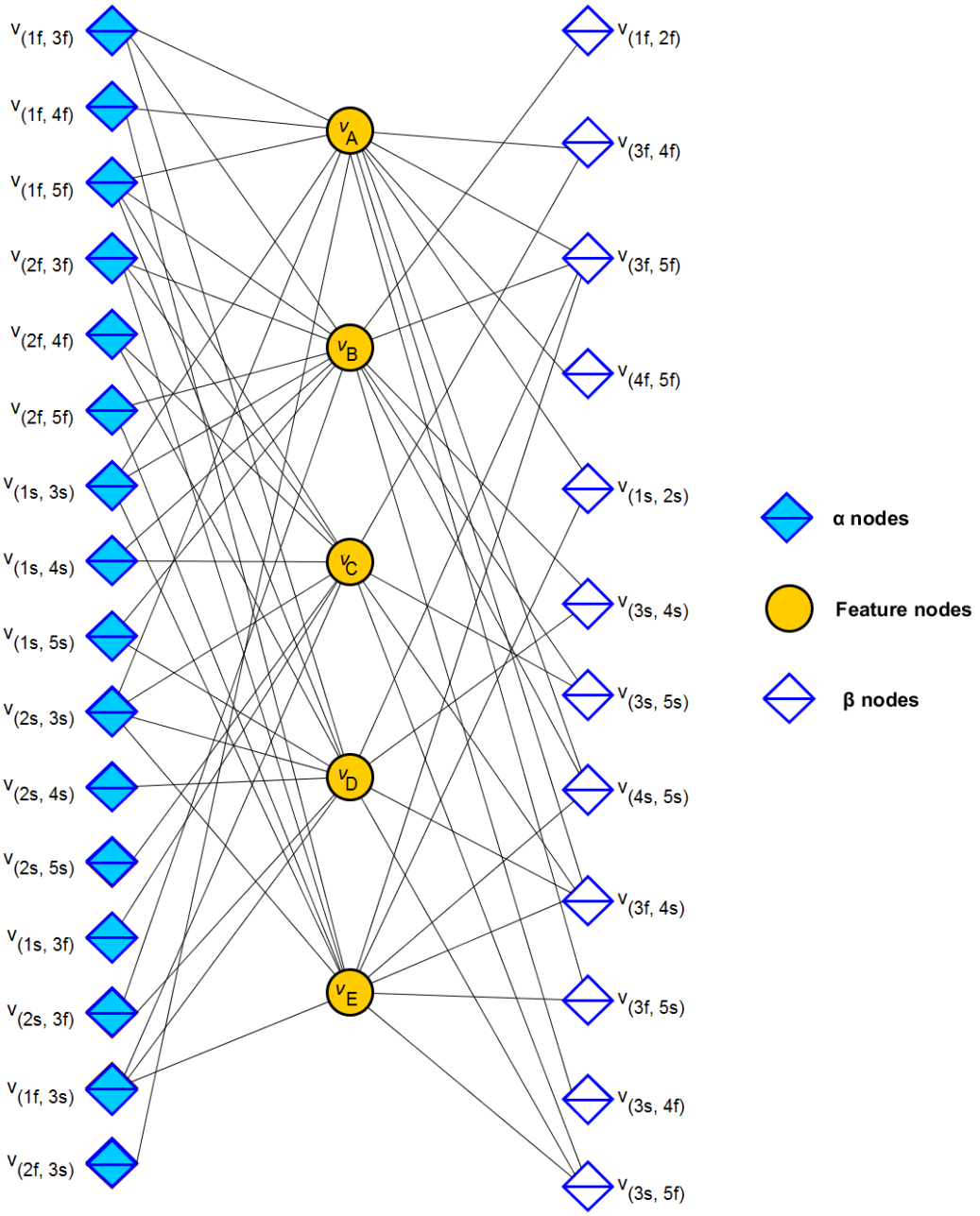
subject to the conditions,

$$\sum_{s=1}^{n} a_{ijs} f_s \geq \alpha \quad \forall (i, j)$$

$$\sum_{s=1}^{n} b_{ijs} f_s \geq \beta \quad \forall (i, j)$$

$$f_s \in \{0, 1\}$$

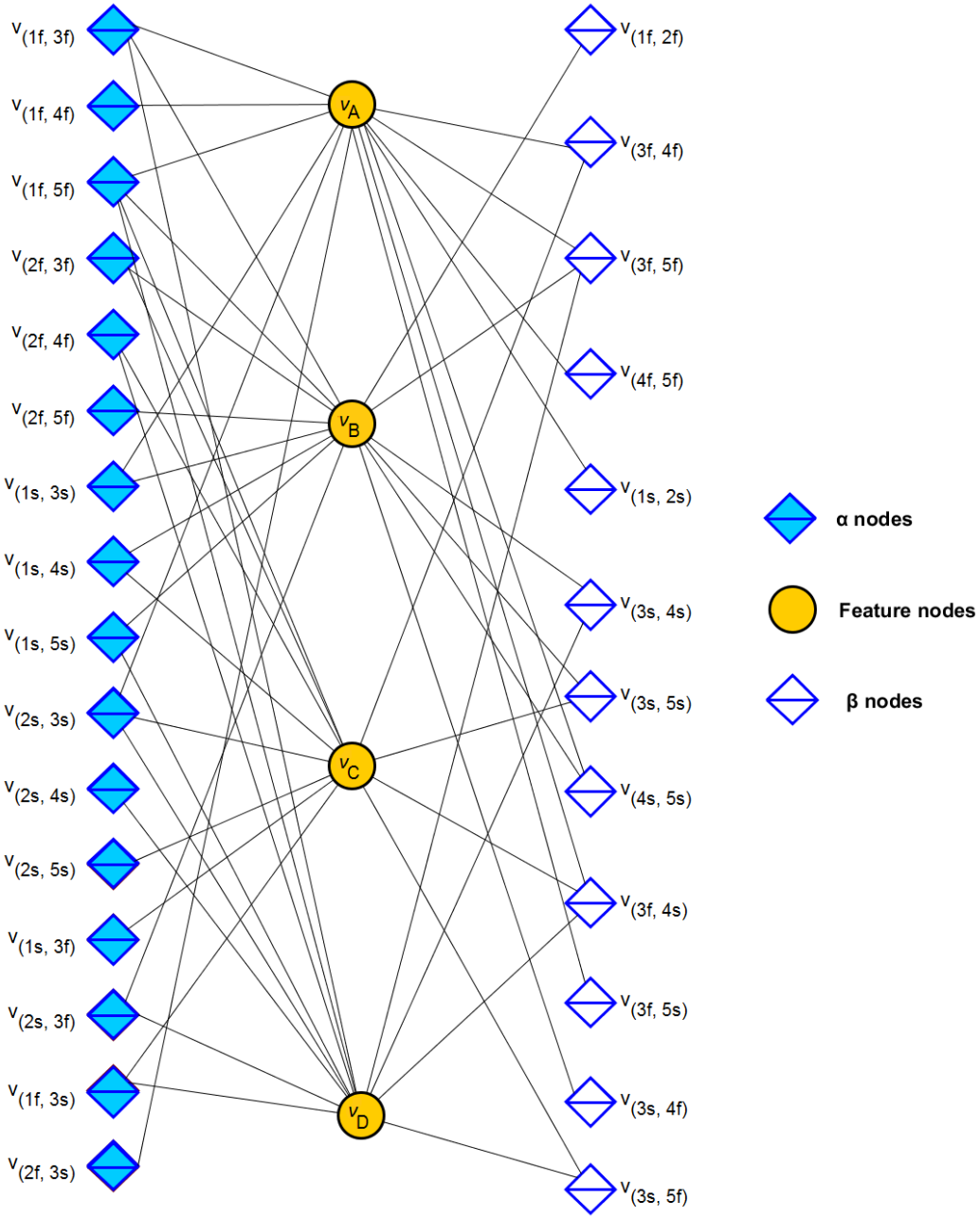
Figure 3.5: Graph representation of the generalised (α, β)-k Feature Set problem created from Table 3.8. Blue nodes represent sample pairs that belong to different classes, α nodes; white nodes represent sample pairs that belong to the same class, β nodes and yellow nodes represent the features.



in which, a_{ijs} and b_{ijs} are the two constraints that indicate the α and β edges that connect feature s and sample pair (i, j) and can be specified as:

$$a_{ijs} = \begin{cases} 1 & if \quad L(i) \neq L(j) \quad \text{and} \quad c_i = c_j \quad \text{and} \quad x_{i,s} \neq x_{j,s} \\ 0 & \text{otherwise} \end{cases}$$

Figure 3.6: An optimal solution for the generalised (α, β)-k Feature Set problem instance from Figure 3.5 with $\alpha = 1$ and $\beta = 1$.



$$b_{ijs} = \begin{cases} 1 & if \quad L(i) = L(j) \quad \text{and} \quad c_i = c_j \quad \text{and} \quad x_{i,s} = x_{j,s} \\ 0 & \text{otherwise} \end{cases}$$

where, $L(i)$ and $L(j)$ are the class labels of samples i and j; c_i and c_j are the colour of the dataset to which the samples i and j belong, and $x_{i,s}$ is the expression value for feature s for sample i.

That is, $a_{ijs} = 1$ if feature (gene) s has different gene expression levels in the sample i and j which belong to different classes but have same colour; and 0 otherwise. Similarly, $b_{ijs} = 1$ if feature (gene) s has the same expression levels in each sample pair (i, j) with the same class label; and 0 otherwise.

The Coloured (α, β)-k Feature Set problem approach attempts to set α as big as possible, k as small as possible and β also as big as possible, in this order of importance. That is, the approach finds the minimum number of features that represent the whole dataset. Any feature set that satisfies the conditions given above is a feasible solution to the Coloured (α, β)-k Feature Set problem. Finally, if more than one solution satisfies all the given conditions, the approach selects the set of features that are undefined. This suggests that the order of the problem's parameters is important. In this method the variables are set sequentially with a four stage process in which each parameter is defined by solving and fixing a sub-problem before moving on to the next.

For this purpose, four optimisation problems are solved and features obtained in the last step are used as the solution because they are considered the most robust and useful of the features compared to other set of features. The four stage approach used to determine the Coloured (α, β)-k Feature Set problem parameters is described below.

(a) Define α^*, the maximum value of α such that there exists an optimal solution for the Coloured (α, β)-k Feature Set problem.

$$\alpha^* = \min(\sum_{s=1}^{n} a_{ijs}),\ \forall (i,j) \in \mathcal{A}$$

(b) Determine k^*, the minimum number of features that are necessary to distinguish the samples that belongs to different classes, considering that at least α^* features do so for each pair of samples.

$$k^* = \min \left(\sum_{s=1}^{n} f_s \right)$$

subject to,

$$\sum_{s=1}^{n} f_s a_{ijs} \geq \alpha^*,\ \forall (i,j) \in \mathcal{A}$$

(c) Determine β^*, the maximum value of β such that exactly k^* features are selected to explain the differences between sample pairs that belong to different classes, and at least α^* features for each pair of α samples, that is by using the values of k^* and α^* obtained in the previous steps. This step

helps to maximises the internal consistency between samples that belong to the same class and provides a much more robust solution.

$$\beta^* = \max \sum_{s=1}^{n} b_{ijs},\ \forall (i,j) \in \mathcal{B}$$

subject to,

$$\sum_{s=1}^{n} f_s = k^*$$

$$\sum_{s=1}^{n} f_s a_{ijs} \geq \alpha^*, \forall (i,j) \in \mathcal{A}$$

(d) Finally find k features that provide the maximum amount of evidence,measured as total cover, by considering the value of α^*, β^* and k^*, obtained in the previous steps, that is,

$$\max \sum_{s=1}^{n} f_s \sum_{i,j=1}^{m} (a_{ijs} + b_{ijs})$$

subject to the conditions;

$$\sum_{s=1}^{n} f_s = k^*$$

$$\sum_{s=1}^{n} f_s a_{ijs} \geq \alpha^*,\ \forall (i,j) \in \mathcal{A}$$

$$\sum_{s=1}^{n} f_s b_{ijs} \geq \beta^*,\ \forall (i,j) \in \mathcal{B}$$

All the formulations presented above are implemented in the C++ tool, which uses IBM's Mixed Integer Linear Optimisation Problem Optimiser CPLEX (http://www-01.ibm.com/software/commerce/optimization/cplex-optimizer/index.html). Details of the input file format for the Coloured (α, β)-k Feature Set problem and the Generalised (α, β)-k Feature Set problem approach are given in Subsection 8.1.2.

Notes

The presented steps represent just one of several possible approaches to solving the problem. Further, a number of variants can be added to these formulations for experimental consideration: for example, weights. Weights can be assigned

to features, samples and edges. The purpose of adding weights is to select features according to their importance in the entire dataset. That is, a low weight implies less importance for that feature. For all these variants it is possible to reformulate the objective function such that the 'coverage' (the final step in Section 3.5) takes weights into account. With or without weights, coverage can be for the α side only, the β side only, or the $\alpha + \beta$ side.

In all these formulations, the samples have been presented as an array of $m \times (n + 1)$ binary values, although this is not strictly necessary. The class labels can be variables taking values over a (typically small) set of classes. The features can have values of any kind, as long as there exists a meaningful comparison test that is able to decide if any two values are to be considered the same or different.

Note that the data for the Coloured (α, β)-k Feature Set problem approach must be discrete, although values in datasets, including microarray datasets, are typically real numbers. To deal with this issue, Fayyad and Irani's entropy-based heuristic [75] can be used on the dataset to remove uninformative features. This univariate selection mechanism is a pre-processing step related to the minimum description length (MDL) principle [75]. The purpose of using this step in this method is twofold: it removes features that are not significantly different between healthy and disease samples (thus it helps by reducing the dimensionality of the problem), and it helps discretise the values (which in turn facilitates the combinatorial approach). For a detailed description of Fayyad and Irani's entropy-based heuristic, please refer to the original paper [75].

In the case of Coloured (α, β)-k Feature Set problem approach, sample pairs of different (α) or same (β) classes are only created within same color.

If C is the number of colours/datasets, M_k, $k = 1, ..., C$ is the number of samples per datasets, $L = c, d$ is the class labels (control and disease), F is the number of features (common to all colours), $M_{k,c}, M_{k,d}$ is the number of samples in dataset k that belongs to class *control, disease* respectively, $\alpha_{mx,k}$ is the Alpha max for the k-th individual (α, β)-k Feature set problem (minimum degree of an α node in the individual k-th (α, β)-k Feature set problem.), $d^*_{\alpha,k}, d^*_{\beta,k}$ is the Maximum degrees to α or β side for the individual k-th (α, β)-k feature set problem, then the number of α nodes is,

$$N_\alpha = \sum_{k=1}^{C} M_{k,c}.M_{k,d} \tag{3.1}$$

and the number of β nodes is,

$$N_\beta = \frac{1}{2}\sum_{k=1}^{c}((M_{k,c}(M_{k,c}-1)+M_{k,d}(M_{k,d}-1)) \tag{3.2}$$

The graph for the Coloured (α, β)-k Feature set problem instance has then $N = N_\alpha + N_\beta + F$ nodes.

If $M^* = \max(m_{k,j})$ is the largest number of samples from any class across all colours,

The upper bound is,

$$N_\alpha \leq C.M^{*2} \tag{3.3}$$

$$N_\beta < C.M^{*2} \tag{3.4}$$

therefore

$$N < 2CM^{*2} + F \tag{3.5}$$

The bounds on the number of nodes shows that the size of the graph grows linearly with the number of colours, and quadratically with the number of samples in a single colour.

To demonstrate the application of the methods described in this chapter, microarray datasets for prostate cancer and AD are analysed in Chapters 4 and 5.

3.6 Conclusion

The proposed methods were explained in this chapter, including details and formulations. The main difficulty with this approach is the computational complexity of this family of problems. The proposed method is as complex as the (α, β)-k Feature Set problem, which is both NP-complete [57, 83] and W[2]-complete [49, 45]. Previous applications of the (α, β)-k Feature Set problem approach reported in the literature have provided interesting results within a reasonable time. The proposed method is likely to be of use in the case of meta-analysis, as the number of newly emerging datasets is increasing. Such datasets may contain in the order of millions of features and tens of thousands of samples.

Chapter 4

Application of the Coloured (α, β)-k Feature Set problem approach to prostate cancer datasets

In this chapter, the application of the Coloured (α, β)-k Feature Set problem in analysis of prostate cancer datasets is explained in detail. This research was published in PLOS One with the title 'A New Combinatorial Optimization Approach for Integrated Feature Selection Using Different Datasets: A Prostate Cancer Transcriptomic Study' [284]. A brief explanation of cancer and prostate cancer is provided in Section 4.1. The datasets are described in detail in Section 4.2. The individual analysis of each dataset is explained in section 4.3, analysis using the (α, β)-k Feature Set problem approach is described in (Subsection 4.3.1), and t-tests, in (Subsection 4.3.2). The integration method for combining the selected datasets is then described in Subsection 4.4 and the combined analysis is explained in Subsection 4.4.2. The results of a comparative analysis with RankProd are presented in Section 4.5 and a sensitivity analysis to check the robustness of the findings is described in Section 4.6. The functional and pathway analyses of the genes identified through this process are presented in Section 4.7. The genes are discussed in the context of the literature in Section 4.8. Finally a conclusion for the chapter is presented in Section4.9.

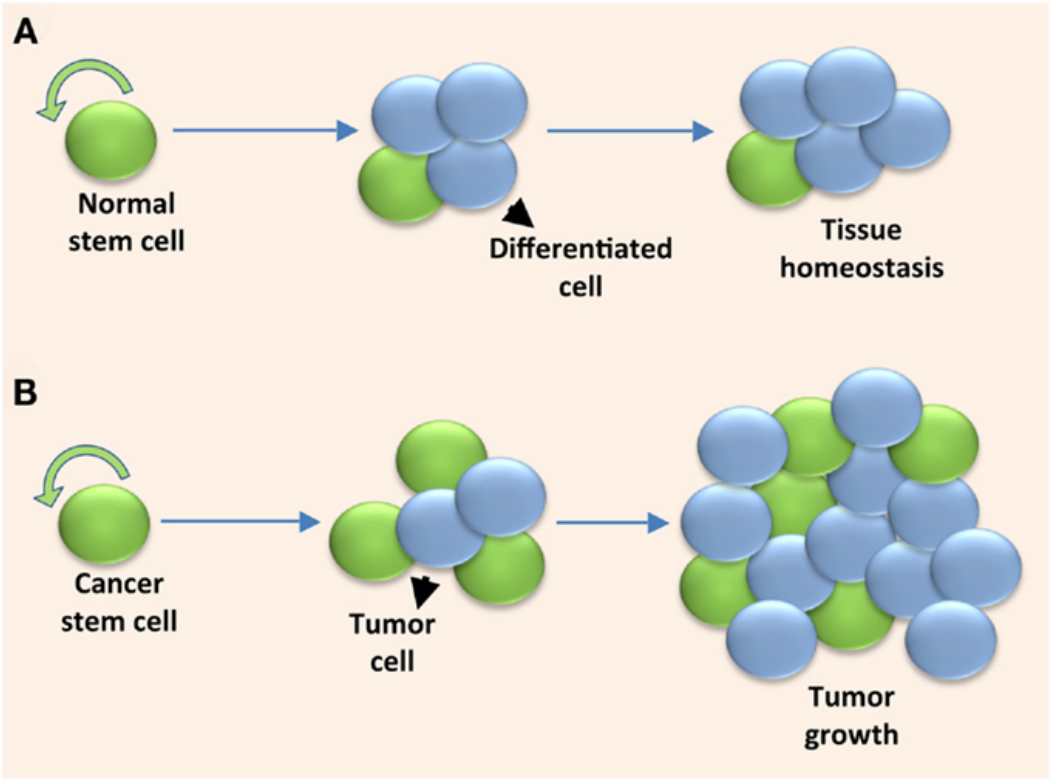
4.1 Cancer

Cancer is a condition in which normal cells reproduce in an uncontrollable manner and begin to invade the surrounding tissues, eventually moving to other parts of the body. Normal cells transform into a cluster of proliferating cancer cells and form a tumour neoplasm, as illustrated in Figure 4.1. A benign tumour can be easily treated and would be considered harmful only if it spread to other parts of the body or continued its growth and endangered the life of the individual.

A malignant tumour is capable of spawning into new tumours by spreading from the primary site. This process is known as metastasis and frequently results in mortality because the metastatic cancer cells prevent normal tissue from performing its vital functions.

Carcinomas are cancers arising from epithelial cells and are the most common type. The two main groups of cancers are sarcoma, which develops from muscle and leukaemia, which develops from blood tissue [162].

Figure 4.1: Cancer stem cells undergoing excessive symmetric cell division. **(A)** A normal cell maintains a controlled series of symmetric and asymmetric cell divisions.**(B)** A cell undergoes excessive symmetric cell divisions due to dysregulated self-renewal, and forms a tumour. *The figure is adapted from [282] with permission.*



The main cause of cancer is changes to genes, which are the unit of information in every cell. These changes, or mutations, cause the cells to grow out of control and form cancer. Genes known as oncogenes have the ability to stimulate cell proliferation. Another type of genes have the ability to suppress cell division. Some of these tumour suppressor genes are involved in DNA repair systems. Multiple mutations in these two groups of genes can transform normal cells into cancer cells [373].

Six hallmarks of cancer were identified by Hanahan and Weinberg [127] in 2000. These six essential alterations are:

- self-sufficiency in growth signals
- insensitivity to growth-inhibitory (antigrowth) signals
- evasion of programmed cell death (apoptosis)
- limitless replicative potential
- sustained angiogenesis
- tissue invasion and metastasis

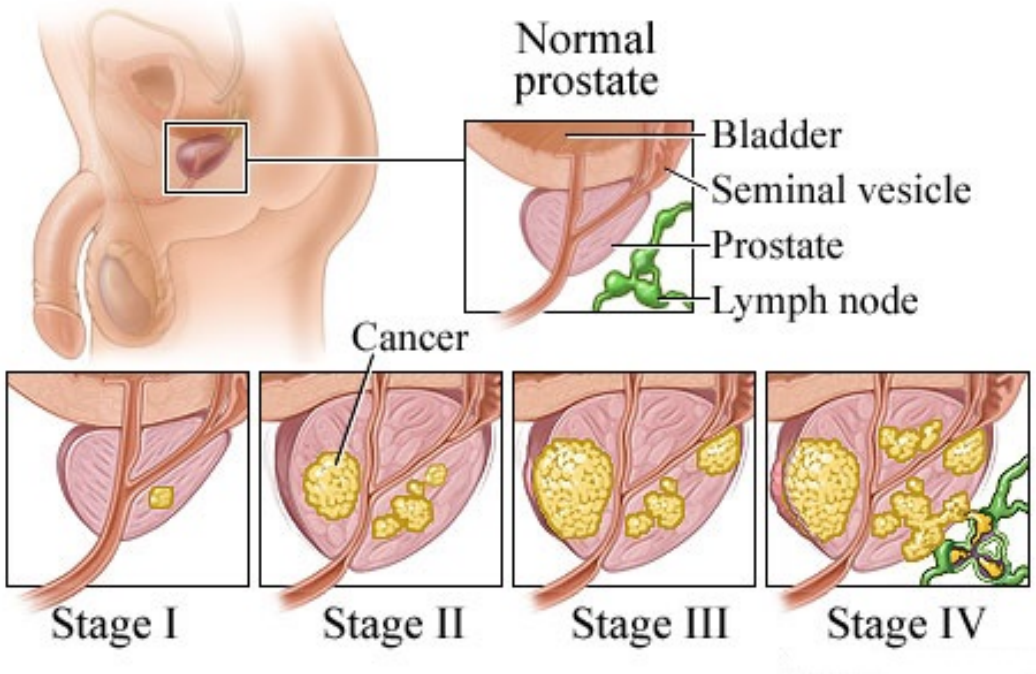
Although there are different types of cancer, prostate cancer datasets are used here for the application of the proposed method. The next section contains a brief explanation of prostate cancer.

4.1.1 Prostate Cancer

Prostate cancer is one of the most common cancers among males, accounting for 10% of cancer-related deaths [364]. Prostate cancer often develops slowly, and a low-grade cancer may not need treatment. The prostate gland plays a key role in the male reproductive system by producing semen, and helping to maintain seminal fluid. Prostate tumours can block urinary flow as the prostate is located on the floor of the pelvis, and surrounds the urinary bladder and urethra. Age, diet and family history are the major risk factors in prostate cancer. Common methods used to diagnose prostate cancer are the digital rectal exam and prostate-specific antigen test. If cancer is suspected, a biopsy is performed via the rectum to obtain tissue samples from the prostate. These samples are examined under a microscope to determine the presence of cancerous cells, and their appearance is typically categorised using Gleason grading. In this way a Gleason score is assigned; cancers with a high score are considered more aggressive Figure 4.2.

Figure 4.2: Prostate cancer development; showing different stages of prostate cancer. *The figure is adapted from [67]with permission*

.



4.2 Datasets

The performance of the Coloured (α, β)-k Feature Set problem approach is now illustrated via a challenging meta-analysis task involving six publicly available prostate cancer microarray datasets. The results are then compared to those obtained using the popular meta-analysis tool RankProd and to the outcomes of analysing each individual dataset by statistical and combinatorial methods alone. The six pre-processed prostate cancer gene expression datasets selected for this study were collected from GEO [14] or from the source given in the original paper. Basic details of all the datasets are summarised in Table 4.1, please refer the original paper for more details on the datasets.

Table 4.1: Summary of prostate cancer datasets used in this study

Name	Platform	Series	NS	Norm	PT	Met	Probes	EF
Singh [325]	Affymetrix [HG-U95Av2]	GSE68907	102	50	52	0	12558	1519
Welsh [374]	Affymetrix [HG-U95Av2]	GPL95	55	9	25	21	12560	2429
Uma [358]	Affymetrix [HG-U95B]	E-GEOD-6919	80	17	63	0	37691	3484
L-2695 [195]	SHBB	GSE3933	26	9	13	4	44161	4288
L-3044 [195]	SHCQ	GSE3933	41	16	23	2	43009	4082
L-3289 [195]	SHBW	GSE3933	45	16	26	3	43009	4953

Name is the name assigned to the study throughout this chapter. **Platform** is the platform details of the dataset, in which the first three are the different versions of Affymetrix platform [213], SHBB, SHCQ and SHBW are the microarray chip manufactured by Stanford school of medicine (http://genome-www5.stanford.edu/) using spotted DNA/cDNA technology. **Series** is the Gene Expression Omnibus Series identifier for the dataset. **Norm** is the number of healthy tissue samples. **PT** is the number of primary tumour samples. **Met** is the number of metastasis samples present in each dataset. **Probes** is the number of probes present in each dataset. **EF** is the number of probes present after entropy filtering.

Each dataset was generated using one of two platforms. The gene expression levels in three studies were measured using custom cDNA two-channel arrays [384] and in the other three, using Affymetrix arrays. These three datasets are named here according to the surname of the first author of the published article. The other three datasets were published by Lapointe et al. [195] Table 4.1 and are given the initial 'L' and the relevant GEO platform number (e.g. L-2695). Singh et al.[325] introduced an outcome prediction model to distinguish between prostate tumour and normal samples. Their dataset is based on 102 tissue samples collected after radical prostatectomy in which 50 were normal samples and 52 were primary prostate cancer samples. The data were generated using Affymetrix HG-U95A v2 (GPL8300) arrays.

The second dataset was contributed by Welsh et al. [374] in 2001. Their study investigated a therapeutic approach to differentiate tumour and normal samples. The dataset results from 55 samples being hybridised to HG-U95A v2 (GPL8300) arrays. The samples were of 25 primary prostate tumour and 9 normal tissues, with the remainder being taken from different donors with different types of cancers.

The third dataset was published by Uma et al. [358] in 2007, based on a study introducing an experimental design to address the differences in cellular content between primary and metastatic prostate tumours. The dataset is based on 63 tumour and 17 normal tissue samples, and was produced using Affymetrix HGU95Av2 arrays.

Lapointe et al. [195] introduced a hierarchical clustering technique to distinguish tumour from normal samples and to identify subclasses of prostate cancer in 2004. Their study was performed using three datasets arising from analysis of cDNA two-channel arrays: L-2695 is based on 26 samples (13 primary tumour tissue, 9 normal tissue and 4 metastasis tissue samples; L-3044 has a total sample count of 41 arising from 23 primary tumour samples, 16 normal samples and 2 metastasis samples; L-3289 is based on 45 samples, of which 26 were primary tumour, 16 normal and 3 metastasis samples.

The present study restricts its focus to the analysis of samples that originated from either primary tumours or normal tissue. The total number of samples is therefore 319, of which 202 were from primary tumours and 117 were from normal tissue.

4.3 Individual dataset analysis

Each individual dataset was analysed using the (α, β)-k Feature Set problem approach and t-test. The individual dataset analysis results were compared

and the genes on common to each were identified. The application of each method and its results are explained below.

4.3.1 Application of the (α, β)-k Feature Set problem approach

As the Coloured (α, β)-k Feature Set problem approach (Chapter 3) proposed in this work is a generalisation of the (α, β)-k Feature Set problem approach for probe set selection, the most natural comparison is to evaluate the Coloured (α, β)-k Feature Set problem approach results by comparing them with those obtained by applying the (α, β)-k Feature Set problem approach to each dataset individually. This means that the feature set problem must be solved for each dataset to identify individual gene signatures that discriminate the sample classes. For that purpose, the (α, β)-k Feature Set problem approach was applied to all six datasets and the results compared.

The application of the (α, β)-k Feature Set problem approach consists of an entropy pre-filtering step and the solution of a combinatorial optimisation problem. The pre-filtering selects features based on the class information content and discards less informative features, thus reducing dimensionality for the subsequent combinatorial problem. Details of the approach are provided in Chapter 2. The characteristics of the individual dataset results are given in Table 4.2.

Table 4.2: The results of the numerical solution to the (α, β)-k Feature Set problem approach for each of the six individual datasets.

Dataset	**Feat.No**	**After EF**	α	β	**Size** (k)
Singh	12558	1519	215	329	754
Welsh	12560	2429	1188	1068	1768
Uma	37691	3484	881	1079	1857
L-2695	44161	4288	2266	2421	3533
L-3044	43009	4028	966	862	1800
L-3289	43009	4953	1397	1216	2696

Dataset is the short name used in this thesis for the dataset. **Feat. No** is the initial number of features (probes) present in the dataset. **After EF** is the number of features remaining after the application of entropy filtering. α and β are the values for the parameters $\alpha_{maximum}$ and $\beta_{maximal}$ for any feasible solution, please refer Chapter 3 for more details. **Size** (k) is the number of features in the resulting solution (signature size) to the individual (α, β)-k Feature Set problem approach for each dataset.

Each dataset produced a molecular signature with a large number of genes. Surprisingly, the number of genes in common between them was only seven, which represents a negligible overlap in the results arising from the experiments. This demonstrates the need for an integrative method: it would be impossible to generate any form of statistical support to link these genes to putative pathways that could be deregulated. However on the positive side, all seven genes have already been reported as having an association with prostate cancer (see Table 4.3).

Table 4.3: List of common genes among all the individual dataset results from Table 4.2

Gene Symbol	Gene Name	Reference
EEF2	Eukaryotic Translation Elongation Factor 2	[268, 357]
SPG20	Spastic Paraplegia 20	No associated reference
ERG	Erythroblastosis Virus E26 Oncogene Homolog	[286, 367]
AMACR	Alpha-Methylacyl-CoA Racemase	[7, 19, 302]
SOX4	SRY (Sex determining Region Y)-box 4	[193, 367]
APOC1	Apolipoprotein C-I	[181, 393]
GUCY1A3	Guanylate Cyclase 1, soluble, alpha 3	[56]

Gene Symbol is the official gene symbols. **Gene Name** is the expanded gene name. **Reference** is the literature reference for each gene which shows the relation with prostate cancer.

4.3.2 Application of t-test

To benchmark the proposed approach against traditional statistical methods, a t-test analysis is performed on each of the individual datasets. The t-test is a statistical significance test method and is used here to select genes that exhibit differential gene expression under two different conditions [230] – in the present case, in normal cells vs. primary tumours – above a certain p-value level of confidence. The procedure for t-tests is as follows.

Let G_1 and G_2 be the mean values of expression of a particular gene in two different class labels 1 and 2, of sizes m_1 and m_2. The t-statistic for this particular gene is computed as:

$$t = \frac{G_1 - G_2}{X\sqrt{\frac{1}{m_1} + \frac{1}{m_2}}}$$

where X is the pooled sample variance

$$X = \sqrt{\frac{m_1 x_1^2 + m_2 x_2^2}{m_1 + m_2 - 2}}$$

Here x_1^2 and x_2^2 are the variance of replicated observations in each condition and $m_1 + m_2 - 2$ is the number of degrees of freedom. In the present study, the `genefilter` Bioconductor package [97] was used with a chosen p-value of 10^{-4} to perform t-tests.

As we are trying to obtain the differentially expressed genes to compare with the (α, β)-k Feature Set problem approach result. The bioconductor package is used for the t-test to get a significant p-value per gene by setting the technical replicates to 100 times. A t-test was applied to each dataset and the results compared with the individual (α, β)-k Feature Set problem approach results. The t-test results for individual datasets are given in Table 4.4.

Table 4.4: t-test results on individual dataset.

Dataset	Feat.No	Signature Size
Singh	1519	616
Welsh	2429	717
Uma	3484	690
L-2695	4288	286
L-3044	4028	654
L-3289	4953	647

Dataset is the short name used in this thesis for the dataset. **Feat.No** is the number of features (probes) present in the dataset before applying t-test. **Signature Size** is the number of genes/features in the resulting solution for each dataset.

Large numbers of genes were filtered out from each dataset using the t-test approach. Only four genes remained in common to all six experiments, and these are *EPCAM, SOX4, EEF2 and AMACR*, given in Table 4.5.

Table 4.5: List of common genes among all the individual dataset results of the (α, β)-k Feature Set problem approach from Table 4.4

Gene Symbol	Gene Name	Reference
EPCAM	Epithelial cell adhesion molecule	[119, 231, 391, 84] [408, 107, 16, 297, 303]
SOX4	SRY (Sex determining Region Y)-box 4	[193, 367]
EEF2	Eukaryotic Translation Elongation Factor 2	[268, 357]
AMACR	Alpha-Methylacyl-CoA Racemase	[7, 302]

Gene Symbol is the official gene symbols. **Gene Name** is the expanded gene name. **Reference** is the literature reference for each gene which shows the relation with prostate cancer.

This was even fewer genes than was found to be in common to all datasets via the individual (α, β)-k Feature Set problem approach analysis. Although three of the genes identified by t-tests (*SOX4, EEF2 and AMACR*) are among the seven identified by the individual (α, β)-k Feature Set problem approach, the fourth, *EPCAM*, was not in Table 4.3.

The results of these individual data analyses, which identified a very small number of important genes in common to each dataset, highlight the necessity of performing a combined study of these datasets, which is now described.

The method used to integrate the datasets, the application of the proposed Coloured (α, β)-k Feature Set problem approach and the results of these procedures are discussed below.

4.4 Integrated data analysis

The integrated dataset was obtained by integrating the features from each individual dataset at the probe level. The Coloured (α, β)-k Feature Set problem approach was then applied to the integrated dataset to identify features that can discriminate the classes present in the integrated dataset.

4.4.1 Integration method

The direct integration of microarray gene expression data from multiple platforms is, in principle, greatly facilitated when there is commonality between the platforms used. However, different gene expression platforms target genes or transcripts differently depending on the sets of probes used. A duplicate spotted probe may represent the same gene in microarray chips. Also, a single probe may be homologous to (and therefore hybridise with) several different genes (or loci) if the specificity of the probe sequence is not sufficiently high. Probes with poor specificity must be discarded from the preliminary analysis as it is difficult to analyse these multiple genes. In addition, the interpretation of results arising from Gene Ontology or pathway-informed databases could be compromised by this multiple mapping problem. In other words, the annotation of these multiply mapped probes results in a list of genes from which it is difficult to identify the dysregulated gene. In addition to these difficulties, there may also be the problem that one probe targeting different regions of the same gene could be indirectly detecting different abundances of protein isoforms. This many-to-many nature of the mapping problem makes it difficult to take a simplistic approach to the essentially different maps produced by the different platforms due to their choice of probe sets.

In the current study, mapping was done at the probe level. To map the probes across the platforms in Table 4.1 to genes, a simple alignment policy was applied as explained below. We have performed the integration of data very simplistically by ignoring the isoforms and alternative splicing as the microarray experiments can not reliably determine alternative splicing and isoform coverage across the whole genome because of the limited number of probes and inherent ambiguity of assignment. A very simplified example of the integration is explianed using Table 4.6 and Table 4.7. The probes were mapped using the hg19-GRCh37 version of the Genome Browser's table produced by the Genome Reference Consortium to avoid the misnaming and misalignment of genes.

Mapping of probes across different platforms poses a major challenge in integrative analysis- a simple cross-reference of the sequence identifiers across different studies rarely works well due to different probe designs in different microarray platforms. To obtain a relatively large number of probes that could be used in the final integrated dataset, those that satisfy any of three conditions were selected. The conditions were:

- where the probes are targeting the same gene: with the assumption that there may be different probes in same dataset or in different dataset that target the same gene.
- where the targeted sequences are overlapping: with the assumption that there may be different probes that have a targeting sequence that overlaps.
- where the targeted sequences are separated by a distance of at most 1000 base pairs: In terms of the length of whole genome, 1000 base pairs is ignorable distance to check if those probes are targeting the same gene.

The probes from each dataset were mapped to genes and the associated transcription start and end position of the targeted genes was compared according to the conditions mentioned above. Whenever there was a targeted gene in common for different probes from multiple datasets, the different combinations of those probes were considered in the combined dataset. Similarly, if the features (the transcription start and end sequences) had an overlap between them, or were separated by at most 1000 base pairs, the combination of those probes was also selected to be part of the combined dataset.

In Table 4.6 and Table 4.7 the transcription start and end positions of each gene were compared and the conditions were applied in turn. For example, the first probe from each dataset, 1003_s_at and 3345_at satisfied the first

Table 4.6: The details for the probes in dataset 1.

Probe ID	txStart	txEnd	Gene
1003_s_at	72634	74249	DUX4L2
1005_at	15193	15277	MIR4273
1007_s_at	102459	104003	DUX2
1030_s_at	112024	112180	RN5-8S1
1034_at	281384	282054	PPP2R3B-AS1

Probe ID is the probe ids present in dataset 1. **txStart** and **txEnd** is the transcription start and end for the corresponding gene. **Gene** is the annotated gene symbol for each probe.

Table 4.7: The details for the probes in Dataset 2.

Probe ID	txStart	txEnd	Gene
3345_at	95776	97391	DUX4L2
3648_at	105003	124000	MIR3
4152_at	112024	112000	TGIF2LY
5030_s_at	102388	104003	CRLF2
5124_at	231384	232054	PPP2R3B-AS1

Probe ID is the probe ids present in dataset 2. **txStart** and **txEnd** is the transcription start and end for the corresponding gene. **Gene** is the annotated gene symbol for each probe.

condition, that is, both were targeting the same gene, DUX4L2. The combination of these two probes was thus selected to be included in the combined dataset. 1030_*s*_*at* and 4152_*at* satisfied the second condition as the targeted sequences were overlapping. The combination of these two probes was also considered to be included in the combined dataset. In the same way, with respect to 1007_*s*_*at* and 3648_*at*, the targeted sequences are closer together than 1000 base pairs, thus satisfying the third condition. In this way each probe was compared and the combination of probes was selected for the combined dataset.

The data that were pre-processed via entropy filtering using Fayyad and Irani's entropy-based heuristic (refer Chapter 3) were used when combining the datasets. This ensured that the selected probes in each individual study carries some differential expression information with respect to the sample classes, and provided a well-defined discretisation which respect to the individual study conditions. Probes in one platform were matched to probes in another platform based on gene names and genomic positions, as explained in Subsection 4.4.1.

The combined dataset contained 319 samples and 16,157 combined probes. Of these, 1405 contained values for all six datasets and 10,729 for three or more datasets, which was annotated to 1454 unique genes. The number of combined probes covering only one dataset was 3425. This uneven cover of datasets is due to some probes being discarded by entropy filtering as uninformative only in some datasets and not in others. However, the large number of combined probes with values in three or more datasets indicates a good level of coverage after dataset integration.

4.4.2 Application of the Coloured (α, β)-k Feature Set problem approach

The Coloured (α, β)-k Feature Set problem approach (refer Chapter 3) was applied to the prepared combined dataset using model 2 of the Coloured (α, β)-k Feature Set (refer Section 3.5 of Chapter 3). Maximum α and β values of 612 and 776 respectively were obtained along with a list of 3190 combined probes that corresponded to 1788 unique genes. To allow a comparison between the number of genes that cover four or more datasets, the genes that cover four datasets or more were considered, and are given in Table 4.8.

Table 4.8: Result of the Coloured (α, β)-k Feature Set problem approach

Datasets	Probes	Genes
Four or more	2272	327
Five or more	1806	186
Six	792	120

Datasets is the considered number of datasets to find the coverage; four or more refers four datasets or more, five or more refers to five datasets or more and six refers to six datasets in the resulted list of combined probes. **Probes** is the resulted number of features (combined probes) after applying the Coloured (α, β)-k Feature Set problem approach. **Genes** is the number of genes correspond to the number of resulted probes.

A gene ordering algorithm presented in [248] was applied to this set of genes to generate heatmaps that highlight the correlation between the resulting genes; shown in Figure 4.3 for the 186 genes that cover five or more datasets and Figure 4.4 for the 120 genes that cover all six datasets.

When considering the identified genes in common between the t-test results, the (α, β)-k Feature Set problem approach and the Coloured (α, β)-k Feature Set problem approach results individually, there were very few in the case of the t-test and the (α, β)-k Feature Set problem results, but the Coloured (α, β)-k

Figure 4.3: Heatmap for the Coloured (α, β)-k Feature Set problem approach resulted genes that cover five or more datasets. It contains 186 up and down regulated genes (rows). The blocks of greenish blue colour represent the absence of gene values in particular datasets. The first colour bar at the bottom indicates Primary Tumour (blue) and Normal (green) samples. The second colour bar represents each sample group in different colour. L-2695 (blue), L-3044 (red), L-3289 (orange), Welsh (grey), Uma (cyan) and Singh (dark grey).The row names of the heatmap is given in Appendix 8.1.3.The magnified version of the figure is available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4480358/figure/pone.0127702.g001/

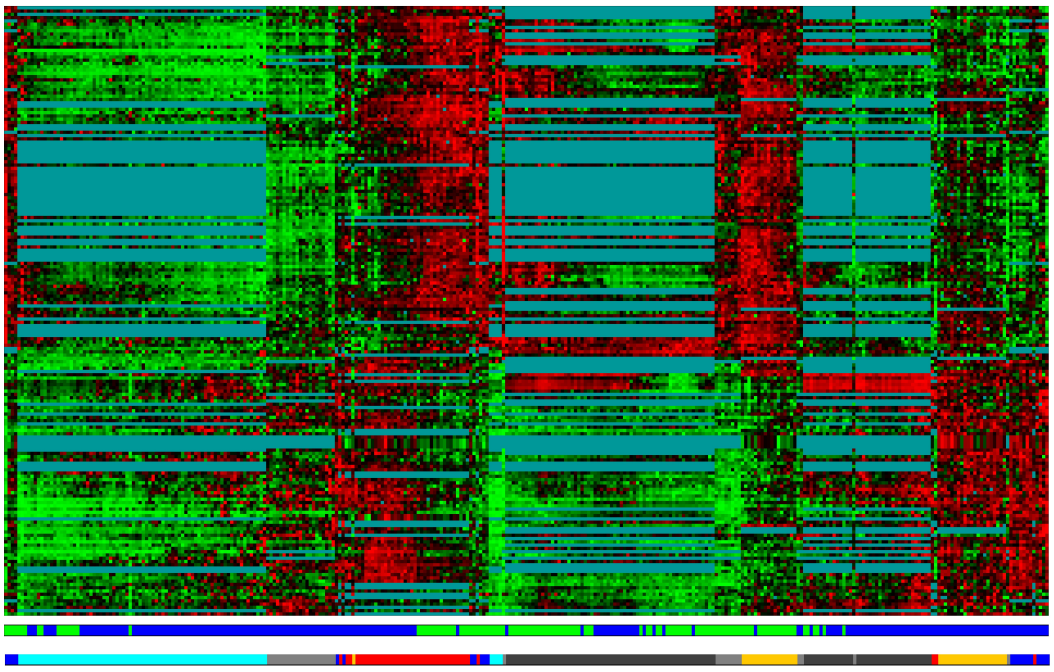
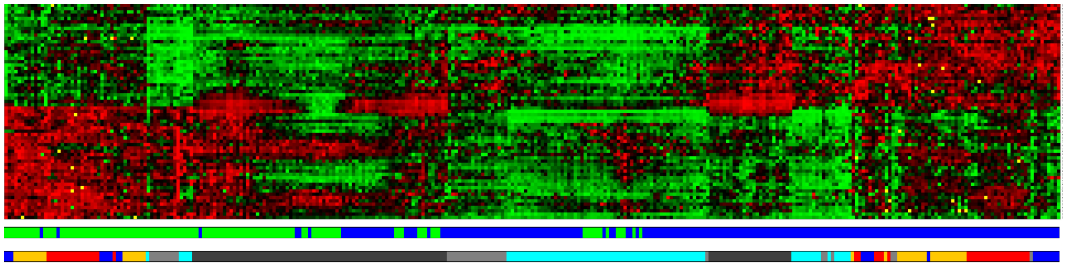


Figure 4.4: Heatmap for the Coloured (α, β)-k Feature Set problem approach resulted 120 genes that cover six datasets. There are 120 up and down regulated genes (rows) which are differentially expressed between normal and tumour classes. The two colour bars at the bottom represent the ordering of samples and sample groups, respectively, as explained in Figure 4.3.The row names of the heatmap is given in Appendix 8.1.3. The magnified version of the figure is available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4480358/figure/pone.0127702.g002/



Feature Set problem approach identified 120 unique genes (see Table 8.2). This represents a substantial difference in the number of common genes identified among the individual datasets. The numbers of overlapping genes identified using the different methods is summarised in Table 4.9.

Table 4.9: Overlapping genes in t-test, the (α, β)-k Feature Set problem approach and the Coloured (α, β)-k Feature Set problem approach.

Datasets	**t-test**	**ABkFS**	**CABkFS**
Six	4	7	120
Five or more	22	57	327
Four or more	36	139	623

Datasets shows the number of datasets considered to find the overlapping, four or more refers four datasets or more, five or more refers to five datasets or more and six refers to six datasets in the resulted list of combined probes. ***t*-test** gives the number of overlapping genes in t-test results for the considered datasets. **ABkFS** gives the number of overlapping genes between individual (α, β)-k Feature Set problem approach result for each case. **CABkFS** gives the number of common genes in the result of the Coloured (α, β)-k Feature Set problem approach result for considered case of datasets.

4.5 Comparison

The results of the current analysis were compared to those obtained using RankProd (see Chapter 2). RankProd ordered the genes by increasing *pfp* (percentage of false positive likelihood) value, and the top genes with a recommended 0.05 *pfp* cut-off from each of the up- and down-regulated gene lists were used for the comparison. This resulted in a list of 1883 genes from the combined dataset (see Table 8.5).

Comparison between the Coloured (α, β)-k Feature Set problem approach result (120 genes) and the RankProd result showed that 80 of the 120 Coloured (α, β)-k Feature Set -identified genes were also present in the top-ranking genes identified by RankProd from the combined dataset. This demonstrates a high level of agreement between the two meta-analysis methods. A summary of the Coloured (α, β)-k Feature Set problem approach and RankProd results is provided in Table 4.10. In addition to the common 80 genes, if the genes that cover four or five datasets in the Coloured (α, β)-k Feature Set problem approach results but were filtered out as non-informative for one or two datasets are considered, the agreement increases to 260 genes out of 327 (almost 80%).

Table 4.10: Comparison of the Coloured (α, β)-k Feature Set problem approach and RankProd result

	RankProd			CABk		
Dataset	**Input**	**pfp**	**Genes**	**6DS (120)**	**5+DS (327)**	**4+DS (623)**
Combined dataset	6929	0.05	1883	80	169	260
		0.01	1484	58	140	214

RankProd is the result of RankProd for Combined dataset with 0.05 and 0.01 *pfp* (percentage of false positive likelihood cut-off). **CABk** is the number of genes resulted from the Coloured (α, β)-k Feature Set problem approach which covered six, five and more, four and more datasets. **6DS** denotes the combination of six datasets. **5+DS** denotes the combination of five or more datasets. **4+DS** denotes the combination of four or more datasets.

However, further analysis with RankProd including genes missing in one or more datasets placed these genes at the top of the list, making further analysis difficult. Similarly, when genes with sparse missing values were included, these genes were artificially escalated in the ranking towards the significant side as more missing values were introduced because the method observed the entire row as a gene and discarded those missing values from the product calculation. This highlights the inability of RankProd to deal with two frequent situations found in microarray datasets.

4.6 Sensitivity analysis

To evaluate the robustness of the proposed method with respect to perturbations in the data, a series of experiments was performed. The presence of noise in gene expression data is difficult to detect, as it depends on platform-specific factors as well as experimental conditions. However, the final manifestation of perturbations in datasets would be a change in the composition of the set of probes that meet the MDL criterion. The robustness of the final integration results was thus analysed with respect to varying compositions of the individual datasets, for different perturbation models, inspired by the leave-one-out approach. Specifically, the following set-ups were modelled: a) removal of one, two and five genes from the combined dataset; and b) removal of one gene from one and two individual datasets. To estimate the worst-case scenario, all genes were restricted to those that appear in the final signature as expressed in all six datasets. In each case, all combined probes corresponding to the chosen gene(s) were removed. An integrated signature was then obtained and compared with the original signature. This procedure was repeated 10 times for the 'a' case and 5 times for the 'b' case, with random selection of gene(s) and dataset(s), and average results reported. Summary results are given in Table 4.11. On average, the results remain the same for more than 97% of the signature list, and the signature sizes remain essentially the same (less than 0.5% increase in the worst case). This points to a highly robust result that does not depend on a (small) set of genes, even if they are in the high coverage set.

4.7 Functional and Pathway Analysis

Functional and pathway analysis has been performed on these 120 genes for further validation of our results. We have used DAVID [148] and STRING [345] for the functional annotation of the association between these genes. Functional annotation of these 120 genes clustered as 8 functionally related groups. Most of the genes in each group are related with prostate cancer and the most known genes in relation with prostate cancer with the clusters of genes can be found in Appendix 8.6.

To identify prostate cancer-related pathways, a pathway analysis was performed using databases such as DAVID [148], KEGG [166] and FatiGO [4]. The pathways identified from DAVID and the associated p-values are provided in Table 4.12. The analysis also identified several other genes that have not

Table 4.11: Result of sensitivity analysis.

	Case a			Case b	
	Exp-1(1 gene)	Exp-2 (2 genes)	Exp-3 (5 genes)	Exp-4 (1 gene / 1DS)	Exp-5 (1 gene / 2DS)
Sig Length	3203 ± 29	3201 ± 28	3205 ± 9	3190 ± 1	3191 ± 1
Overlap	97 ± 3	98 ± 3	98 ± 1	99 ± 1	99 ± 1
New Features	46 ± 11	51 ± 18	77 ± 18	19 ± 17	28 ± 21
Cover of New Features	3.6	3.17	3.4	1.47	1.4
Sig length variation	0.41%	0.36%	0.47%	0.01%	0.03%

Case a is the result of sensitivity analysis after removing one gene **(Exp-1)**, two genes **(Exp-2)** and five genes **(Exp-3)** from the combined dataset. **Case b** gives the result of sensitivity analysis after removing one gene from one **(Exp-4)** and two **(Exp-5)** individual datasets. Values in parenthesis are the standard deviations for the 10 repetitions (Case a) and 5 repetitions (Case b). **Sig Length** is the average signature length in each case. **Overlap** is the average percentage overlap with the original result. **New Features** is the average number of new features in each case. **Cover of New Features** is the average cover of new features. **Sig length variation** is the average signature length variation compared to the original result.

been previously identified in the literature as being associated with prostate cancer.

Table 4.12: The top 14 resulted pathways from pathway analysis.

Pathway name	**Pathway Classification**	**p-value**	**Reference**
Integrin signalling pathway	Cell communication	1.03E-08	[312]
Smooth Muscle Contraction	Organismal Systems; Circulatory system	5.98E-08	[371, 407]
Oxytocin signalling pathway	Organismal Systems; Endocrine system	1.23E-08	[350]
Collagen biosynthesis & modifying enzymes	Metabolism; Amino acid metabolism	1.13E-07	[286, 307]
Axon guidance	Development	1.34E-06	[109]
Gap junction trafficking	Cell communication	3.12E-06	[242]
Protein digestion and absorption	Organismal Systems; Digestive system	3.6E-06	[307]
Ras activation	Regulation of translation and transcription	3.46E-05	[2]
regulation of pgc-1a	Cell motility	3.61E-05	[411]
Assembly of collagen fibrils & other multimeric structures	Metabolism	3.68E-05	[43]
CREB phosphorylation	Metabolism; Energy metabolism	6.31E-05	[147]
Syndecan-1-mediated signalling events	Genetic Information Processing	6.3E-05	[206]
NCAM1 interactions	Signal Transduction	6.72E-05	[400]
regulators of bone mineralization	Metabolism	6.7E-05	

Pathway Name is the name of the pathways resulted from DAVID. **Pathway Classification** is the class of each pathway. **p-value** is the respective p-value for each pathway. **Reference** is the publications that show the relation of each pathway with prostate cancer.

4.8 Discussion

Microarray technology has had a tremendous influence on cancer research in terms of assessing the presence of cancer cells in patient tissues. The rapid acquisition of microarray data makes it possible to integrate this large amount of data across a range of platforms. In this study, robust cancer gene expression signatures common to all datasets were identified. The comparison of the results from the proposed method with individual study results highlighted the advantages of meta-analysis over individual studies. The comparison of the results from the proposed method with those from a method that is considered to be state of the art demonstrates the robustness of this method.

The (α, β)-k Feature Set problem approach results for each individual dataset provided signatures of reasonable size capable of discriminating between primary tumours and normal samples. However, even though the individual signatures consisted of a large number of features, the number of common genes was limited to seven, which is too low for further analysis. The result of Coloured (α, β)-k Feature Set problem approach analysis provided a vastly larger number of genes that should be targeted for further analysis. The combined dataset contains 10,729 of 16,157 combinations of probes from three or more datasets, which confirms that a good coverage of all the datasets was achieved by using the proposed method of integration. Further, the Coloured (α, β)-k Feature Set problem approach results showed that around two thirds (2272 out of 3190) of the identified features were found in four or more datasets.

Even though the (α, β)-k Feature Set problem approach and t-tests provided good results on individual datasets, a large number of genes were eliminated from the common set of genes, which may include important biomarkers. When the data integration was performed, the number of resulting genes was significantly increased relative to just genes common to individual datasets. The proposed method makes it possible to uncover robust biomarkers by increasing the sample size to a sufficient level to achieve statistical significance, and helps to capture consistent features that might have been masked because of limitations of the individual studies. As a reasonable number of genes has been identified, they may also provide more information about prostate cancer.

Application of the Coloured (α, β)-k Feature Set problem approach provided evidence of a high level of agreement with the top-ranked genes in the RankProd result, where almost 80% of the signature was included in RankProd's result. However, the RankProd results were considerably larger in size, hindering interpretation. Additionally, as mentioned previously, RankProd artificially reduces the rank of any gene with missing values (escalating its position to

the significant side of the cut-off point), which i) restricts applicability to the genes represented in all platforms; and ii) introduces non-linear rank scaling in the presence of scattered missing values. In contrast, the Coloured (α, β)-k Feature Set problem approach automatically deals with any amount of missing data (i.e. a gene not present in a given dataset may still be significant in explaining a large number of sample pairs in the other datasets), providing a more reliable result. Although not used in this way in the current investigation, the Coloured (α, β)-k Feature Set problem approach allows for weights to be assigned to genes and samples independently to account for an external perceived relative confidence in each experimental condition, if so desired.

The sensitivity analysis revealed a high level of consistency for the original solution. Each step of the sensitivity analysis confirmed that the results of the proposed method did not rely on a single or a small set of genes. The results of the analysis also showed that the significance of a given gene was not dependent on a single dataset. The consistency of the results indicates the robustness of the proposed method and validates the findings.

It is not surprising that most of the signature genes have been reported to be related with prostate cancer. For instance, *AMACR* [157, 6, 144, 406, 220, 302], *HPN* [94, 120, 177], *SOX4* [368, 193, 246, 309, 128], *DAXX* [192, 355], *EPB41L3* [18, 314, 315], *CXCR3* [72, 186, 260, 261, 323, 372, 382], *TGFB3* [293, 37], *EEF2* [268, 383] are the most well-known biomarkers for prostate cancer. As identified by the Gene Ontology Consortium, most of the signature genes are involved in the cell cycle (*MYH11*), regulation of transcription (*SOX4, SMARCC2, ZIM2, PDLIM5, ZNF217, PSIPI, ACRC, PEG3, TAF1, ZMYM3*), receptor activity (*JAM3, TAPBP, COL4A5, CXCR3, COL4A6, HPN, COL9A2, PTPRN2, COL6A1*) and other biological activities such as transportation, cell adhesion and cell organisation.

Most of the genes mentioned above are highly correlated with prostate cancer. However, only some of them were identified in individual dataset results. Genes were also identified here that participate in the same pathway class as genes related to prostate cancer, but have not previously been reported in relation to prostate cancer. For instance, *NUDT3* has not been mentioned in relation to prostate cancer, although it is known to participate in the 'collagen biosynthesis and modifying enzymes' pathway, which has been identified as a prostate cancer-related pathway in [126]. Together with its identification as a signature gene in the current study, this strongly suggests that this gene may also have some influence on prostate cancer development.

Interestingly, the most significant pathway represented in our results is the integrin signalling pathway and focal adhesion. Integrins are transmembrane

receptors and play an important role in cell survival, proliferation, migration, gene expression and activation of growth factor receptors. Studies show that integrins are down regulated in the transition from normal prostate tissue to primary localised prostate cancer [108]. The list of signature genes identified in the current study includes *COL4A5, COL4A6, COL6A1 and ITGB1BP2* which are known to participate in the integrin signalling pathway.

Smooth muscles found in the walls of the reproductive tracts of males and females are made up of actin and myosin, which together provide the capacity for the muscles to contract and relax. The prostate helps to control urine flow and ejaculation via contraction and relaxation of its smooth muscle layers. The uncontrolled contraction of prostate smooth muscle may result in urinary tract problems as well as prostate growth [312]. The smooth muscle contraction pathway has already been reported to be associated with prostate cancer [337]. The list of signature genes includes *MYH11, MYL6, MYL6B and GUCY1A3*, which are related to smooth muscle contraction.

The collagen biosynthesis pathway is responsible for collagen production. Studies have shown that the Gleason score increases with decreasing cancer collagen content [36]. The list of signature genes includes *TGFB, COL4A5, COL4A6, COL6A1 and COL9A2*, which are associated with collagen biosynthesis.

The outcomes of the current research support the claim that the proposed method is a viable and very useful meta-analysis method for feature selection. The functional and pathway analysis results showed that the Coloured (α, β)-k Feature Set problem approach is capable of uncovering genes with significant and biologically relevant functions that other, non-integrative methods have failed to identify.

4.9 Conclusion

his chapter presented the application of the Coloured (α, β)-k Feature Set problem approach in multi-platform integration analysis without the need for normalisation of data between datasets. The results indicate that the method is capable of providing highly significant signatures, even when individual datasets are small and thus lack informational content. The method is generic and does not depend on inherent properties of gene expression data, allowing it to be potentially applied to any dataset where the notions of features, class-based classification and equality between feature values is meaningful. In applying this methodology to an integrated prostate cancer dataset, this study has identified potential novel prostate cancer-associated pathways and genes. As the

number of cancer datasets increases this novel and robust method we be able to be used to combine more cancer datasets and identify more candidate pathways and genes.

Chapter 5

Application of the Coloured (α, β)-k Feature Set problem approach to Alzheimer's Disease datasets

This chapter applies the Coloured (α, β)-k Feature Set problem approach to AD datasets. This research was published in PLOS One with the title 'Identification of Differentially Expressed Genes Through Integrated Study of Alzheimer's Disease Affected Brain Regions' [285]. A brief explanation of AD and related studies is provided in Section 5.1. The datasets used in this study are explained in detail in Section 5.2. The individual analysis of the data is explained in Subsection 5.2.1, along with the results for each brain region. The integration method for combining the selected datasets is described in Subsection 5.3.1 and the combined analysis is explained in Subsection 5.3.2. The results of the analysis are compared with those of RankProd and GeneMeta in Section 5.4 and the sensitivity analysis to check the robustness of the findings is outlined in Section 5.5. The resulting genes are discussed with related references in Section 5.6. Finally a conclusion for the chapter is provided in Section 5.7.

5.1 Alzheimer's Disease

AD is a progressive and degenerative neurological disorder characterised by the loss of mental ability. It is the most common cause of dementia with loss of cognitive functions and memory. AD kills nerve cells and makes changes in

neurons and neurotransmitters that affect the communication between neurons, leading to brain function loss. The most common clinical features of AD are the aggregation of β-amyloid into plaques; the presence of hyperphosphorylated tau protein in self-assembled tangles and filaments; and the loss of connections between nerve cells in the brain, which leads to brain atrophy [25, 234]. Although the process and development of AD are not understood, it is likely that deterioration of the brain begins well before symptoms occur. The common symptoms of AD are difficulty with remembering recent events, thinking and reasoning, speaking and writing, making judgements and decisions, planning and performing familiar tasks; and changes in personality and behaviour [29]. With the progression of AD, most parts of the brain become seriously damaged and shrink dramatically due to widespread cell death. In the advanced stages of AD, individuals lose their ability to communicate, to recognise family and loved ones, and to care for themselves.

AD is not a part of normal aging, but increasing age is the strongest risk factor in AD. Three in 10 people over the age of 85 and 1 in every 8 people over 65 are expected to develop AD. Family history and genetics, mild cognitive impairment (MCI), past head trauma, lifestyle and heart health, life-long learning and social engagement are all related to the risk of developing AD. The risk of developing AD is higher if a first-degree relative (parent or sibling) has the disease. The genetic mechanism of familial AD remains unknown. People with MCI have a higher chance of developing AD, but this is not inevitable and can be prevented by developing a healthy lifestyle. Some studies show that risk factors in heart disease may also increase the risk of developing AD [280, 82, 390]. Further, there is a relationship between life-long involvement in mentally and socially stimulating activities, and reduced risk of AD [190, 333, 379].

Diagnosis of AD is usually based on the patient's medical history, mental status testing and physical testing. Even though AD is associated with several histopathological markers, such as extracellular β-amyloid plaques and neurofibrillary tangles (NFTs) within neurons [91, 201, 319], these can only be evaluated in the post-mortem brain or in rare surgical circumstances; thus physicians have turned to less invasive methods to diagnose AD, such as neuroimaging. Two available imaging techniques for AD diagnosis are positron emission tomography, which identifies a pattern of reduced glucose with the help of a measure of cerebral glucose metabolic rate [253, 159, 363]; and magnetic resonance imaging, which identifies brain atrophy [89, 159, 238].

These techniques help to identify damage in tissue or vessels in the brain, rather than directly predicting the risk of developing AD. Most research on

AD focusses on the hippocampus (HIP) because it is the first brain region to be affected by AD [255].

However, other brain regions are functionally associated with memory, attention, perceptual awareness, thought, language and consciousness, which are also affected in AD. For example, the entorhinal cortex (EC) region works as a mediator of learning and memory. EC and HIP together play an important role in the visual processing hierarchy and thereby receive signals for object representations [224]. The posterior cingulate (PC) cortex helps with visual perception and memory recollection [199, 200]. The middle temporal gyrus (MTG) is involved in some basic functions such as recognition of faces and ascertaining of distance. [74]. The superior frontal gyrus (SFG) is associated with self-awareness and with the action of the sensory system [27]. The visual cortex (VCX) processes visual information by receiving visual data from the lateral geniculate body of the thalamus [90].

Since the first characterisation of the disease symptoms in 1906 by Dr Alois Alzheimer, genesis of AD has remained elusive. Only in 1993 was the *APOE* (apolipoprotein E) gene found to be associated to AD. Several studies have since identified other DEGs in AD-affected brain regions [316]. However, ELISA (enzyme-linked immunosorbent assay) measurement of β-amyloid, total tau and phospho-tau-181 in cerebrospinal fluid are the most advanced and accepted method for AD diagnosis. It is estimated that by 2050, approximately 80 million people globally will be suffering from AD [149]. Therefore, it is a great and important challenge to find reliable biomarkers to understand the mechanism behind AD [78, 330].

An 18-protein signature in peripheral blood plasma was identified by Ray et al. [292] that can be used to predict the presence of AD before clinical symptoms are evident. That study used a single classifier approach to identify a panel of proteins that helps to decide whether patients with MCI will develop AD in the next two to six years. Soon after, Gomez Ravetti and Moscato [104] used the same dataset and identified a five-protein biomarker, which was a subset of Ray's 18-protein signature, that is sufficient to provide the same result with better accuracy to predict AD. Using Ray's dataset, Paula et al. [277] identified a specific pattern of cell signalling imbalance that can predict AD in patients with MCI.

In 2010, Gomez Ravetti et al. [105] identified a clear pattern of up- and down-regulated genes that was related to the HIP region and reveals alterations in calcium, insulin, phosphatidylinositol and Wnt signalling. They also found that gene probes that are strongly associated with AD severity are involved in synaptic function, neurofilament bundle assembly, and neuronal plasticity and

inflammation. They showed that gene homologues of *EGR1* (early growth response protein 1)-zif268, Egr-1 or *ZENK*-together with other members of the *EGR* family, play an important role in short- and long-term memory and neuronal plasticity in the brain. All these studies were concentrated on data from a specific brain region. Combined studies on different brain regions may provide more information with regard to gene dysregulation driving development of AD pathogenesis.

To this end, in 2008, Liang et al. [209] performed a combined study of post-mortem gene expression data from six different brain regions and identified several DEGs (*APOE, BACE1, STUB1 (CHIP), FYN, GGA1, SORL1*), as well as pinpointing genes with significant expression changes in AD across brain regions. Ray and Zhang [291] and Kim, Basak and Holtzman [179] performed a four-region study to gain knowledge about different regions, and built a co-expression gene network to characterise the similarities and differences among the regions. They also found that the MTG region shows an early AD pathology compared to other regions. A network-based systems biology approach was proposed by Liu et al. [216] to study AD-related pathways and their dysfunctions among six brain regions. This identified the most significant AD-related pathways across the six regions.

Further, Lambert et al. [194] conducted a large-scale, two-stage meta-analysis of genome-wide association studies in 74,046 individuals. They identified 11 new susceptibility loci that are significant in relation to AD. Bertram et al. [22] performed a meta-analysis of AD genetic association studies and identified 20 polymorphisms in 13 genes that are strongly associated with AD. A genome-wide association meta-analysis of neuropathologic features of AD identified nine new loci, involving the genes *ABCA7, BIN1, CASS4, CD33, MEF2C, MS4A6A, PICALM, SORL1, ZCWPW1*, which are significant in regard to AD pathogenesis [15].

As the sample size of individual gene expression microarray datasets is low, computational methods can be used to integrate the data from different microarray studies. Greco et al. [112] proposed an integration method to combine microarray gene expression data from Affymetrix GeneChip experiments to investigate tissue-selective expression patterns. A computational approach was developed by Wang et al. [366] for a genome-wide analysis of human tissue-selective gene expression data from heterogeneous sources.

In short, a combined study of similarities and differences among different AD-affected brain region datasets is expected to provide a better understanding of AD pathogenesis. Most studies have reported a large number of genes, and yet the results are conflicting [118]. Due to the exceedingly large number

of genes related to AD in different brain regions, it has become virtually impossible to systematically follow, evaluate, interpret or compare these findings.

A robust characterisation of the transcriptomic risk factors related to AD requires an integrated study. Here, a combined analysis is performed using gene expression data from six AD-affected brain regions from the well-known Liang gene expression dataset [210]. The dataset contains data for the EC, HIP, MTG, PC, SFG and VCX regions and is re-analysed in this study to identify common genes among six AD-affected brain regions, using the Coloured (α, β)-k Feature Set problem approach [284]. As a robust feature selection method, the Coloured (α, β)-k Feature Set problem approach can handle the integrated dataset to find the minimum set of genes that is common and significant to explain AD across regions. Also, individual region analyses are performed to identify region-specific genes and compare them with the common genes. A functional and pathway analysis for the identified genes that are strongly associated with AD development is also performed.

5.2 Datasets

This study used a publicly available Affymetrix microarray gene expression dataset for AD contributed by Liang et al. [210], deposited in GEO[14] under the series number GSE5281. The dataset consists of post-mortem data from 161 samples, 74 of which were from non-demented controls and 87 from AD patients, with a mean age of 79.8 ± 9.1 years. The samples had been collected (with a mean post-mortem interval of 2.5 hours) from three AD centres following the deaths of clinically and neuropathologically classified AD-affected individuals. The samples were taken from six different brain regions: EC, HIP, MTG, PC, SFG and VCX. Details of the samples in each region are provided in Table 5.1.

In these datasets, some samples are of different brain regions from the same individual; however this information was not fully available in the original publication. One of the goals of the current study was to verify the robustness of the obtained signatures, accounting for inter-individual variability. The accompanying clinical information was used to identify from which brain region each sample was taken. The clinical data indicate that there are overlapping and repeating samples between regions. The details of the samples are given in Table 8.7.

As a pre-processing step, Fayyad and Irani's entropy-based heuristic was applied to the data from each region to remove uninformative features (I refer the reader to Chapter 3). The filter searches for a discretisation threshold (or

Table 5.1: Sample details that belongs to different regions.

Region	Control	Affected	Total
EC	13	9	23
HIP	13	10	23
MTG	12	16	28
PC	12	9	22
SFG	9	23	34
VCX	12	19	31

Region is the name of different regions in the data, that is,EC - Entorhinal Cortex, HIP - Hippocampus, MTG - Middle temporal gyrus, PC - Posterior cingulate cortex, SFG - Superior frontal gyrus, VCX - visual cortex. **Control** is the number of controls in each region. **Affected** is the number of diseased samples in the data. **Total** is the total number of samples in each region.

possibly a set of thresholds) maximising the class entropy gain. As different tissues have different gene expression profiles, the filter was applied to each region separately. This method helps to remove features that are not significantly different in control and AD samples and to reduce the dimensionality of the problem. It also facilitates the combinatorial approach by discretising the values of features.

5.2.1 Individual Data Analysis

For each region the (α, β)-k Feature Set problem approach was applied, giving a region-specific signature. As explained in Chapter 3, this approach provided a significant set of genes that collectively maximised the inter-class discrimination and intra-class coherency [20, 46]. The method helps to select a minimum set of features that collectively provide the maximum amount of evidence to differentiate the control and AD samples in each brain region. The resulting probes were annotated using BioMart [326] and pathway analysis was performed using EASE [142].

Features that are differentially expressed were identified by applying the (α, β)-k Feature Set problem approach [20, 46] (I refer the reader to Chapter 3) to data from each region separately. DEGs and specific dysregulated pathways together provide new insights into the pathogenesis of AD. The DEGs and related pathways for each region are analysed and explained in the following sections.

The purpose of this study was to identify significant genes in each region associated with the presence of AD, to identify common genes among all the

regions, and to bring together all these results with previous studies to develop a sketch of region specificity in relation to AD. Hence, to analyse the top listed genes with high statistical relevance from the results, EASE was used to analyse the resulting genes and obtain a corresponding EASE score – a modified version of Fisher's exact p-value used for gene-enrichment analysis – to identify dysregulated pathways.

An EASE score $\leqslant$ of 0.06 indicates that the gene is specifically associated with the pathway in the context of the list provided to EASE.

Further, to simplify the functional and pathway analysis, Bonferroni correction was applied. This is a conservative adjustment to the EASE score to control for any multiple comparison effects. It was applied to the resulting list of features from individual and combined analysis to select features with a Bonferroni-corrected p-value (BF-value) < 0.0001. The top 15 features according to BF-value were then discussed further in relation to AD.

TThe analysis of each region resulted in a long list of genes that are significantly associated with AD. Genes that had a p-value (BF-value) < 0.0001 were selected for functional and pathway analysis.

The number of resulting genes before and after Bonferroni correction for each region is given in Table 5.2.

A gene ordering methodology used previously [248] was applied to the resulting set of genes for each region to generate a heatmap that highlights associations among the genes.

We also find some probes corresponding to the microRNA precursors for each region. We must notice that the gene expression microarray platform used in this study, Affymetrix HGU133 plus v2.0, is only capable of detecting microRNA precursors and not mature microRNA sequences. The mention of these precursors is however relevant, as they are a necessary for the synthesis of functional mature sequences.

The individual analysis result of each region data is given in the following sections.

Entorhinal Cortex(EC):

EC is the main channel between HIP and the neocortex and is involved in the long-term cognitive memory formation [87]. In particular, EC supplies information to HIP from multiple senses and translates information to neocortex with the help of a neurotransmitter called glutamate. Studies have already been shown that EC is one of the region affected by AD in the early stage itself [28, 102, 329, 175].

Table 5.2: Significant genes in different brain regions.

Region	EF-Probes	Result	Genes	$BF_{.0001}$	Pathways
EC	11504	4558	3762	108	24
HIP	11501	7779	5594	475	55
MTG	12607	6398	4941	1138	81
PC	15907	12690	7402	206	21
SFG	8785	5473	4344	47	23
VCX	5332	2185	1900	11	0

Region is the acronym of the different brain regions in the data. **EF-Probes** is the number of features that pass Fayyad-Irani's entropy filter for each region. **Result** is the number of features resulted from the (α, β)-k Feature Set approach for each region (k, signature size). **Genes** is the number of genes obtained by annotating the resultant signature. $\boldsymbol{BF_{.0001}}$ is the number of genes that are used for further analysis with a Bonferroni corrected p-value $<$0.0001. **Pathways** is the number of related pathways by annotating the genes with **BF $<$ 0.0001**.

In EC, we found 4558 probes differentially expressed of which 108 have a BF-value $<$ 0.0001 mapping to 94 genes and 24 related pathways, given in Table 5.3. Among the list of resulted features we identified 10 microRNAs that are differentially expressed in EC region. The details of microRNAs can be found in Table 5.4.

Hippocampus (HIP):

HIP is a part of the temporal lobe that is absolutely necessary for forming new memories. It is common that AD affects the HIP early and severely before affecting any other part of the cortex [255], which shows memory is the first thing that starts to get falter in AD. Several studies shows that APOE plays a prominent role in HIP damage through impaired blood flow and the consequent lack of oxygen [313, 222, 327, 227]. In HIP, we identified 7779 probes that are differentially expressed, of which 475 have a BF-value $<$ 0.0001, mapping to 392 genes and 55 significant pathways with EASE score $<$ 0.06. The list of pathways are given in Table 5.5. From the resulted list of features, we find 12 differentially expressed microRNAs in HIP. The details of microRNAs are given in Table 5.6.

The Middle Temporal Gyrus (MTG):

MTG is a gyrus on the temporal lobe of the brain which is involved in a number of cognitive processes such as semantic memory, language processing

Table 5.3: List of dysregulated pathways related with EC region

Name	**EASE score**	**Gene List**
Porphyrin and chlorophyll metabolism	1.16E-03	*EPRS; HMOX1*
Metabolism of Cofactors and Vitamins	1.49E-03	*EPRS; HMOX1; PTPRA*
Cell cycle	4.62E-03	*CDKN2C; PTPRA*
Cell Growth and Death	5.62E-03	*CDKN2C; PTPRA*
Transcription	1.33E-02	*TAF5L*
Replication and Repair	1.40E-02	*POLH*
Translation	1.47E-02	*EPRS*
gamma-Hexachlorocyclohexane degradation	1.55E-02	*PTPRA*
Nucleotide Metabolism	1.55E-02	*POLH*
Riboflavin metabolism	1.62E-02	*PTPRA*
Purine metabolism	1.64E-02	*POLH*
Apoptosis	1.65E-02	*PTPRA*
Biodegradation of Xenobiotics	1.77E-02	*PTPRA*
Pyrimidine metabolism	1.77E-02	*POLH*
aminoacyl-tRNA biosynthesis	1.77E-02	*EPRS*
DNA polymerase	1.77E-02	*POLH*
Phosphatidylinositol signaling system	1.79E-02	*PTPRA*
Sorting and Degradation	1.85E-02	*SRPR*
Arginine and proline metabolism	1.88E-02	*EPRS*
Glutamate metabolism	1.94E-02	*EPRS*
Signal Transduction	1.96E-02	*PTPRA*
Protein export	1.97E-02	*SRPR*
Amino Acid Metabolism	1.98E-02	*EPRS*
Basal transcription factors	1.98E-02	*TAF5L*

Name is the name of pathways. **EASE Score** is the pathway score by annotating using EASE. **Gene List** is the list of genes involved in the pathway.

Table 5.4: List of microRNAs resulted from EC region.

Probe	Gene-id	Description	Gene-name	Gene-source	mirbase-id
228528_at	ENSG00000275668		AL035209.1	MI0000107	hsa-mir-29b-2
230127_at	ENSG00000207870	microRNA 221 [Source:HGNC Symbol;Acc:HGNC:31601]	MIR221	MI0000298	hsa-mir-221
1557293_at	ENSG00000207650	microRNA 570 [Source:HGNC Symbol;Acc:HGNC:32826]	MIR570	MI0003577	hsa-mir-570
239269_at	ENSG00000273834	microRNA 612 [Source:HGNC Symbol;Acc:HGNC:32868]	MIR612	MI0003625	hsa-mir-612
201639_s_at	ENSG00000216133	microRNA 939 [Source:HGNC Symbol;Acc:HGNC:33682]	MIR939	MI0005761	hsa-mir-939
33132_at	ENSG00000216133	microRNA 939 [Source:HGNC Symbol;Acc:HGNC:33682]	MIR939	MI0005761	hsa-mir-939
224739_at	ENSG00000276753	microRNA 6821 [Source:HGNC Symbol;Acc:HGNC:49980]	MIR6821	MI0022666	hsa-mir-6821
218695_at	ENSG00000276472	microRNA 6847 [Source:HGNC Symbol;Acc:HGNC:50022]	MIR6847	MI0022693	hsa-mir-6847
203669_s_at	ENSG00000276987	microRNA 6848 [Source:HGNC Symbol;Acc:HGNC:50176]	MIR6848	MI0022694	hsa-mir-6848
203189_s_at	ENSG00000277703	microRNA 7113 [Source:HGNC Symbol;Acc:HGNC:49947]	MIR7113	MI0022964	hsa-mir-7113

Probe is the respective probe ids resulted from EC. **Gene-id** is the ensembl id of the genes. **Description** gives the details about the genes. **Gene-name** is the respective gene names. **Gene-source** is the external gene source for each gene. **mirbase-id** is the respective microRNA id for each probe-id.

Table 5.5: List of dysregulated pathways related with HIP region

Pathway Name	**EASE score**	**Gene List**
Glyoxylate and dicarboxylate metabolism	6.24E-05	GRHPR; MDH1; MDH2; MTHFD1
Pyruvate metabolism	6.23E-04	GRHPR; MDH1; MDH2; PDHA1
Carbohydrate Metabolism	1.08E-03	GRHPR; IDH3A; MDH1; MDH2; MTHFD1; PDHA1; TALDO1
Proteasome	1.30E-03	PSMD4
Citrate cycle (TCA cycle)	1.63E-03	IDH3A; MDH1; MDH2
Valine, leucine and isoleucine biosynthesis	2.05E-03	BCAT1; PDHA1
Energy Metabolism	2.55E-03	GLUD1; MDH1; MDH2; NDUFA6; NDUFB1; NDUFV2
Reductive carboxylate cycle (CO2fixation)	2.64E-03	MDH1; MDH2
Glutamate metabolism	2.79E-03	GLUD1
Ubiquitinmediated proteolysis	3.03E-03	CDC16; SMURF2; UBE2D3
DNApolymerase	3.19E-03	POLG; REV3L
Replication and Repair	3.19E-03	POLG; REV3L
Carbon fixation	4.16E-03	MDH1; MDH2
Cell cycle	4.62E-03	ACP1; ATR; DUSP8; PRKDC; RB1
Sorting and Degradation	4.65E-03	CDC16; PSMD4; SMURF2; UBE2D3
Purine metabolism	4.85E-03	ADCY3; PDE5A; POLG; REV3L
Cell Growth and Death	5.31E-03	ACP1; ATR; DUSP8; PRKDC; RB1
Butanoate metabolism	5.39E-03	PDHA1
Nucleotide Metabolism	6.50E-03	ADCY3; PDE5A; POLG; REV3L
Oxidative phosphorylation	6.86E-03	NDUFA6; NDUFB1; NDUFV2
Lysine degradation	6.88E-03	DOT1L
Metabolism of Cofactors and Vitamins	7.10E-03	ACP1; BCAT1; DUSP8; MTHFD1
Pyrimidine metabolism	7.12E-03	POLG; REV3L
Neurodegenerative Disorders	7.79E-03	CREBBP; GNB1; SMURF2
Metabolism of Other Amino Acids	7.90E-03	CSAD; GLUD1
Amino Acid Metabolism	8.98E-03	BCAT1; DOT1L; GLUD1; PDHA1

Name is the name of the pathways. **EASE Score** is the pathway score by annotating using EASE. **Gene List** is the genes that belongs to each pathway.

List of dysregulated pathways related with HIP region, continuation

Pathway Name	EASE score	Gene List
Riboflavin metabolism	9.07E-03	ACP1; DUSP8
Apoptosis	9.07E-03	ACP1; DUSP8
Signal Transduction	9.37E-03	ACP1; DUSP8; EGFR
Folate biosynthesis	9.48E-03	MTHFD1
Biodegradation of Xenobiotics	9.68E-03	ACP1; DUSP8
Phosphatidylinositol signaling system	9.77E-03	ACP1; DUSP8
One carbon pool byfolate	1.10E-02	MTHFD1
Transcription	1.36E-02	POLR1D
Urea cycle and metabolism ofamino groups	1.51E-02	GLUD1
Metabolism of Complex Lipids	1.77E-02	UGCG
Sphingoglycolipid metabolism	1.91E-02	UGCG
Glycolysis / Gluconeogenesis	2.32E-02	PDHA1
Valine, leucine and isoleucine degradation	2.73E-02	BCAT1
Cell Communication	3.14E-02	PDPK1
Dentatorubropallidoluysian atrophy (DRPLA)	3.55E-02	SMURF2
Arginine and proline metabolism	3.96E-02	GLUD1
Taurine and hypotaurine metabolism	4.37E-02	CSAD
Metabolism of Complex Carbohydrates	4.78E-02	RPN1
Huntington's disease	5.00E-02	CREBBP
Integrin-mediated cell adhesion	5.07E-02	RPN1
Pentose phosphate pathway	5.19E-02	TALDO1
RNA polymerase	5.23E-02	POLR1D
Nitrogen metabolism	5.41E-02	GLUD1
Pantothenate andCoA biosynthesis	5.59E-02	BCAT1
Alzheimer's disease	5.64E-02	GNB1
D-Glutamine and D-glutamate metabolism	5.75E-02	GLUD1
MAPK signaling pathway	5.82E-02	EGFR
N-Glycansbiosynthesis	5.86E-02	RPN1
gamma-Hexachlorocyclohexane degradation	5.97E-02	RPN1

Name is the name of the pathways. **EASE Score** is the pathway score by annotating using EASE. **Gene List** is the genes that belongs to each pathway.

Table 5.6: List of microRNAs resulted from HIP region.

Probe	Gene_id	Description	Gene_name	Gene-source	mirbase_id
204355_at	ENSG00000221585	microRNA 1226 [Source:HGNC Symbol;Acc:HGNC:33922]	MIR1226	MI0006313	hsa-mir-1226
212674_s_at	ENSG00000221585	microRNA 1226 [Source:HGNC Symbol;Acc:HGNC:33922]	MIR1226	MI0006313	hsa-mir-1226
224598_at	ENSG00000221394	microRNA 1229 [Source:HGNC Symbol;Acc:HGNC:33924]	MIR1229	MI0006319	hsa-mir-1229
231061_at	ENSG00000199169	microRNA 367 [Source:HGNC Symbol;Acc:HGNC:31781]	MIR367	MI0000775	hsa-mir-367
204881_s_at	ENSG00000266315	microRNA 4668 [Source:HGNC Symbol;Acc:HGNC:41545]	MIR4668	MI0017298	hsa-mir-4668
227858_at	ENSG00000266041	microRNA 4690 [Source:HGNC Symbol;Acc:HGNC:41707]	MIR4690	MI0017323	hsa-mir-4690
213156_at	ENSG00000207770	microRNA 568 [Source:HGNC Symbol;Acc:HGNC:32824]	MIR568	MI0003574	hsa-mir-568
1557293_at	ENSG00000207650	microRNA 570 [Source:HGNC Symbol;Acc:HGNC:32826]	MIR570	MI0003577	hsa-mir-570
224739_at	ENSG00000276753	microRNA 6821 [Source:HGNC Symbol;Acc:HGNC:49980]	MIR6821	MI0022666	hsa-mir-6821
222029_x_at	ENSG00000275010	microRNA 6834 [Source:HGNC Symbol;Acc:HGNC:50108]	MIR6834	MI0022679	hsa-mir-6834
218695_at	ENSG00000276472	microRNA 6847 [Source:HGNC Symbol;Acc:HGNC:50022]	MIR6847	MI0022693	hsa-mir-6847
200775_s_at	ENSG00000207603	microRNA 7-1 [Source:HGNC Symbol;Acc:HGNC:31638]	MIR7-1	MI0000263	hsa-mir-7-1

Probe is the respective probe ids resulted from EC. **Gene-id** is the ensembl id of the genes. **Description** gives the details about the genes. **Gene-name** is the respective gene names. **Gene-source** is the external gene source for each gene. **mirbase-id** is the respective microRNA id for each probe-id.

and integration of information from different senses [74]. Many studies have shown the active neuronal loss for AD in the MTG region of the brain [129, 44, 110]. 6398 differentially expressed features have been identified in relation with this region of which 1138 have a BF-value < 0.0001, mapping to 1020 genes and 20 significant pathways with EASE score < 0.06. The list of pathways can be found in Table 5.7. From the resulting list of features, we find 13 microRNAs that are differentially expressed in MTG. The details of microRNAs are given in Table 5.8.

The Posterior Cingulate Cortex (PC):

The PC is part of the cingulate cortex, which is a highly connected and metabolically active brain region that is functionally involved in learning and spatial memory. Studies have identified amyloid deposition and reduced metabolism in PC in the progress of AD, and this brain region is also significantly reduced in size in AD patients compared with controls [161, 199, 200, 135]. The analysis performed here identified 12,690 differentially expressed features, of which 206 had a BF-value < 0.0001, mapping to 187 genes and 21 significant pathways with an EASE score < 0.06, pathways are in Table 5.9 The resulting list of features includes 22 microRNAs that are differentially expressed in PC, given in Table 5.10.

The Superior Frontal Gyrus (SFG):

SFG is located at the superior part of the prefrontal cortex and it makes up about one third of the frontal lobe. Stimulation and activation of SFG is involved in self awareness and plays a role in working memory as well as manipulation of this memories to accomplish cognitive tasks like planning for the future, judgement, decision-making skills, attention span, and inhibition. Damage in SFG can cause in problems performing these functions [339, 27]. Several studies have been identified the presence of frontal hypo metabolism in relation with AD [258, 361]. We have identified 5473 differentially expressed features in SFG region of which 47 have a BF-value < 0.0001, mapping to 42 genes and 23 significant pathways with EASE score < 0.06. The pathways are listed in Table 5.11. Among the resulted list of features, we find 13 microRNAs that are differentially expressed in SFG, given in Table 5.12.

Table 5.7: List of dysregulated pathways related with MTG region

Pathway Name	EASE score	Gene List
ATP synthesis	1.24E-04	ATP5C1; ATP5G1; ATP5G3; ATP5J2; ATP5O; ATP6V1B2; ATP6V1D; ATP6V1E1; ATP6V1G2; ATP6V1H
Neurodegenerative Disorders	1.95E-04	ALS2; APP; CLTA; CREBBP; GNG12; GNG3; GSK3B; NEFH; NEFL; PPP3CA; SMURF2; SNCA; UBE2G2; UBE2L3; UCHL1
Pentose and glucuronateinter conversions	3.18E-04	UGP2
Amyotrophic lateralsclerosis (ALS)	1.30E-03	ALS2; NEFH; NEFL; PPP3CA
Parkinson's disease	1.30E-03	SNCA; UBE2G2; UBE2L3; UCHL1
Energy Metabolism	1.33E-03	ASNS; ATP5C1; ATP5G1; ATP5G3; ATP5J2; ATP5O; ATP6V1B2; ATP6V1D; ATP6V1E1; ATP6V1G2; ATP6V1H; COX6C; FH; GLS; GOT2; MDH1; MDH2; ME3; NDUFA1; NDUFA6; UQCRH
Alzheimer's disease	1.65E-03	APP; GNG12; GNG3; GSK3B; SNCA
Cell Communication	1.95E-03	ARHGEF7; MAPK6; PIK3R1; RAC1; SEPP1; SORBS1; TLN2
Integrin-mediated cell adhesion	1.95E-03	ARHGEF7; MAPK6; PIK3R1; RAC1; SEPP1; SORBS1; TLN2
Sorting and Degradation	2.07E-03	ANAPC5; CUL1; CUL3; PSMA1; PSMB2; PSMD1; PSMD14; PSMD4; SEC61A2; SMURF2
Pyruvate metabolism	2.29E-03	ACACB; MDH1; MDH2; ME3; PDHA1; PDHB
Carbon fixation	2.58E-03	GOT2; MDH1; MDH2; ME3
Lysine degradation	2.71E-03	PLOD2
Glutamate metabolism	3.27E-03	ABAT; EPRS; GLS; GOT2
Proteasome	3.29E-03	PSMA1; PSMB2; PSMD1; PSMD14; PSMD4
Tyrosine metabolism	3.35E-03	GOT2
Oxidative phosphorylation	3.45E-03	ATP5C1; ATP5G1; ATP5G3; ATP5J2; ATP5O; ATP6V1B2; ATP6V1D; ATP6V1E1; ATP6V1G2; ATP6V1H; COX6C; NDUFA1; NDUFA6; UQCRH
Prostaglandin andleukotriene metabolism	3.99E-03	LTA4H; YWHAZ
Reductive carboxylate cycle (CO2fixation)	4.05E-03	FH; MDH1; MDH2
Alanine and aspartate metabolism	4.05E-03	ABAT; ASNS; GOT2
Valine, leucine and isoleucine biosynthesis	4.05E-03	IARS; PDHA1; PDHB

Name is the name of the pathways. **EASE Score** is the pathway score by annotating using EASE. **Gene List** is the genes that belongs to each pathway.

List of dysregulated pathways related with MTG region, continuation

Pathway Name	EASE score	Gene List
Carbohydrate Metabolism	5.29E-03	ABAT; ACACB; ALDH6A1; FH; MDH1; MDH2; ME3; PDHA1; PDHB; PRPS1; SUCLG1; UGP2
Riboflavin metabolism	5.54E-03	ACP1; PTPN3; PTPRD; PTPRN2; PTPRR
Inositol phosphate metabolism	5.74E-03	ITPKB
Ubiquitin mediated proteolysis	5.88E-03	ANAPC5; CUL1; CUL3; SMURF2
Propanoate metabolism	5.88E-03	ABAT; ACACB; ALDH6A1; SUCLG1
Signal Transduction	5.98E-03	ACP1; CDS1; DGKE; ITPKB; MAPK10; MAPK9; PTPN3; PTPRD; PTPRN2; PTPRR
Butanoate metabolism	6.37E-03	ABAT; PDHA1; PDHB
Sphingoglycolipid metabolism	6.38E-03	UGCG
Nitrogen metabolism	6.40E-03	ASNS; GLS
Phenylalanine, tyrosine andtryptophan biosynthesis	6.40E-03	GOT2; YARS
Basal transcription factors	6.40E-03	GTF2B; TAF9
Citrate cycle (TCA cycle)	6.92E-03	FH; MDH1; MDH2; SUCLG1
gamma-Hexachlorocyclohexane degradation	6.94E-03	ACP1; PTPN3; PTPRD; PTPRN2; PTPRR
Phosphatidylinositol signaling system	7.09E-03	ACP1; CDS1; DGKE; ITPKB; PTPN3; PTPRD; PTPRN2; PTPRR
Glyoxylate and dicarboxylate metabolism	7.21E-03	MDH1; MDH2
Glycerolipid metabolism	7.33E-03	CDS1; DGKE; GNPAT; PAFAH1B1; YWHAZ
Cell cycle	7.66E-03	ACP1; GSK3B; PTPN3; PTPRD; PTPRN2; PTPRR; RB1
Apoptosis	7.68E-03	ACP1; PTPN3; PTPRD; PTPRN2; PTPRR
Biosynthesis of Secondary Metabolites	7.85E-03	ACACB; GOT2
MAPKsignaling pathway	7.85E-03	MAPK10; MAPK9
Transcription	7.85E-03	GTF2B; TAF9
aminoacyl-tRNA biosynthesis	7.92E-03	EPRS; IARS; YARS
Purine metabolism	8.00E-03	ADCY1; GDA; GUCY1B3; HPRT1; PRPS1
Amino Acid Metabolism	8.02E-03	ABAT; AHCYL1; ALDH6A1; ASNS; EPRS; GLS; GOT2; IARS; PDHA1; PDHB; PLOD2; YARS
Metabolism of Complex Lipids	8.17E-03	CDS1; DGKE; GNPAT; ITPKB; LTA4H; PAFAH1B1; UGCG; YWHAZ
Nucleotide Metabolism	8.22E-03	ADCY1; GDA; GUCY1B3; HPRT1; PRPS1; UGP2
Arginine and proline metabolism	8.34E-03	EPRS; GOT2

Name is the name of the pathways. **EASE Score** is the pathway score by annotating using EASE. **Gene List** is the genes that belongs to each pathway.

List of dysregulated pathways related with MTG region, continuation

Pathway Name	EASE score	Gene List
Metabolism of Cofactors and Vitamins	8.61E-03	ACP1; EPRS; NMNAT2; PTPN3; PTPRD; PTPRN2; PTPRR
Huntington's disease	8.72E-03	CLTA; CREBBP
Cysteine metabolism	8.77E-03	GOT2
Cell Growth and Death	8.79E-03	ACP1; GSK3B; PTPN3; PTPRD; PTPRN2; PTPRR; RB1
Metabolism of Other Amino Acids	9.26E-03	ABAT; AHCYL1; GLS
Biodegradation of Xenobiotics	9.37E-03	ACP1; PTPN3; PTPRD; PTPRN2; PTPRR
Alkaloidbiosynthesis I	9.41E-03	GOT2
Glycolysis / Gluconeogenesis	9.55E-03	PDHA1; PDHB
Metabolism of Complex Carbohydrates	9.75E-03	CMAS; FUCA1; UGP2
Translation	9.92E-03	EPRS; IARS; RPL15; RPS28; YARS
Phenylalanine metabolism	1.18E-02	GOT2
Ribosome	1.24E-02	RPL15; RPS28
Selenoamino acid metabolism	1.48E-02	AHCYL1
Lipid Metabolism	1.55E-02	ACACB; STS
Porphyrin and chlorophyll metabolism	1.79E-02	EPRS
Inositol metabolism	2.09E-02	ALDH6A1
Androgen and estrogen metabolism	2.39E-02	STS
Methionine metabolism	2.70E-02	AHCYL1
Nucleotide sugars metabolism	3.00E-02	UGP2
Galactose metabolism	3.30E-02	UGP2
Phospholipid degradation	3.60E-02	YWHAZ
Pentose phosphate pathway	3.91E-02	PRPS1
Tetracycline biosynthesis	4.21E-02	ACACB
Protein export	4.51E-02	SEC61A2
Valine, leucine and isoleucine degradation	4.82E-02	ALDH6A1
Starch and sucrose metabolism	5.03E-02	UGP2
Dentatorubropallidoluysian atrophy (DRPLA)	5.12E-02	SMURF2
N-Glycan degradation	5.24E-02	FUCA1
Aminosugars metabolism	5.33E-02	CMAS
Nicotinate and nicotinamide metabolism	5.42E-02	NMNAT2
beta-Alanine metabolism	5.63E-02	ABAT
	5.73E-02	ACACB
D-Glutamine and D-glutamate metabolism	5.94E-02	GLS

Name is the name of the pathways. **EASE Score** is the pathway score by annotating using EASE. **Gene List** is the genes that belongs to each pathway.

Table 5.8: List of microRNAs resulted from MTG region.

Probe-id	Gene-id	Description	Gene-name	Gene-source	mirbase-id
226448_at	ENSG00000221290	microRNA 1182 [Source:HGNC Symbol;Acc:HGNC:35263]	MIR1182	MI0006275	hsa-mir-1182
224598_at	ENSG00000221394	microRNA 1229 [Source:HGNC Symbol;Acc:HGNC:33924]	MIR1229	MI0006319	hsa-mir-1229
237217_at	ENSG00000264638	microRNA 3152 [Source:HGNC Symbol;Acc:HGNC:38379]	MIR3152	MI0014179	hsa-mir-3152
208750_s_at	ENSG00000264944	microRNA 3620 [Source:HGNC Symbol;Acc:HGNC:38917]	MIR3620	MI0016011	hsa-mir-3620
229926_at	ENSG00000265452	microRNA 3682 [Source:HGNC Symbol;Acc:HGNC:38916]	MIR3682	MI0016083	hsa-mir-3682
212593_s_at	ENSG00000265827	microRNA 4680 [Source:HGNC Symbol;Acc:HGNC:41541]	MIR4680	MI0017312	hsa-mir-4680
202365_at	ENSG00000265455	microRNA 4700 [Source:HGNC Symbol;Acc:HGNC:41806]	MIR4700	MI0017333	hsa-mir-4700
203281_s_at	ENSG00000263506	microRNA 5193 [Source:HGNC Symbol;Acc:HGNC:43534]	MIR5193	MI0018172	hsa-mir-5193
213156_at	ENSG00000207770	microRNA 568 [Source:HGNC Symbol;Acc:HGNC:32824]	MIR568	MI0003574	hsa-mir-568
1557293_at	ENSG00000207650	microRNA 570 [Source:HGNC Symbol;Acc:HGNC:32826]	MIR570	MI0003577	hsa-mir-570
217770_at	ENSG00000273555	microRNA 6812 [Source:HGNC Symbol;Acc:HGNC:50116]	MIR6812	MI0022657	hsa-mir-6812
224739_at	ENSG00000276753	microRNA 6821 [Source:HGNC Symbol;Acc:HGNC:49980]	MIR6821	MI0022666	hsa-mir-6821
209140_x_at	ENSG00000277402	microRNA 6891 [Source:HGNC Symbol;Acc:HGNC:50243]	MIR6891	MI0022738	hsa-mir-6891

Probe is the respective probe ids resulted from EC. **Gene-id** is the ensembl id of the genes. **Description** gives the details about the genes. **Gene-name** is the respective gene names. **Gene-source** is the external gene source for each gene. **mirbase-id** is the respective microRNA id for each probe-id.

Table 5.9: List of dysregulated pathways related with PC region

Pathway Name	**EASE score**	**Gene List**
Sorting and Degradation	2.27E-04	ANAPC5; CDC16; CUL1; CUL3; PSMB1; PSMD12
Ubiquitin mediated proteolysis	2.32E-04	ANAPC5; CDC16; CUL1; CUL3
Oxidative phosphorylation	2.80E-04	ATP5A1; ATP5F1; ATP5O; COX4I1; COX5B; UQCRC2
Energy Metabolism	1.34E-03	ATP5A1; ATP5F1; ATP5O; COX4I1; COX5B; UQCRC2
ATP synthesis	1.43E-03	ATP5A1; ATP5F1; ATP5O
Ribosome	3.03E-03	RPL14; RPL15; RPL29; RPL4
Translation	4.17E-03	RPL14; RPL15; RPL29; RPL4
Proteasome	5.61E-03	PSMB1; PSMD12
Glycerolipid metabolism	6.73E-03	AGPAT1; LIPA
Lipid Metabolism	7.40E-03	HSD17B7; LIPA
Metabolism of Complex Lipids	9.12E-03	AGPAT1; LIPA
Porphyrin andchlorophyll metabolism	1.37E-02	BLVRA
Metabolism of Cofactors and Vitamins	1.46E-02	BLVRA
Glutathione metabolism	1.55E-02	GPX3
Androgen and estrogen metabolism	1.64E-02	HSD17B7
Metabolism of Complex Carbohydrates	1.73E-02	EXT2
Huntington's disease	1.82E-02	CLTA
Metabolism of Other Amino Acids	1.91E-02	GPX3
Bile acid biosynthesis	2.01E-02	LIPA
Chondroitin /Heparan sulfate biosynthesis	2.10E-02	EXT2
Neurodegenerative Disorders	2.19E-02	CLTA

Name is the name of the pathways. **EASE Score** is the pathway score by annotating using EASE. **Gene List** is the genes that belongs to each pathway.

Table 5.10: List of microRNAs resulted from PC region.

Probe-id	Gene-id	Description	Gene-name	Gene-source	mirbase-id
232642_at	ENSG00000221120	microRNA 1224 [Source:HGNC Symbol;Acc:HGNC:33923]	MIR1224	MI0003764	hsa-mir-1224
204355_at	ENSG00000221585	microRNA 1226 [Source:HGNC Symbol;Acc:HGNC:33922]	MIR1226	MI0006313	hsa-mir-1226
212674_s_at	ENSG00000221585	microRNA 1226 [Source:HGNC Symbol;Acc:HGNC:33922]	MIR1226	MI0006313	hsa-mir-1226
224598_at	ENSG00000221394	microRNA 1229 [Source:HGNC Symbol;Acc:HGNC:33924]	MIR1229	MI0006319	hsa-mir-1229
226871_s_at	ENSG00000221410	microRNA 1238 [Source:HGNC Symbol;Acc:HGNC:33933]	MIR1238	MI0006328	hsa-mir-1238
229437_at	ENSG00000275402	microRNA 155 [Source:HGNC Symbol;Acc:HGNC:31542]	MIR155	MI0000681	hsa-mir-155
226144_at	ENSG00000223244	microRNA 1909 [Source:HGNC Symbol;Acc:HGNC:35393]	MIR1909	MI0008330	hsa-mir-1909
200033_at	ENSG00000265695	microRNA 3064 [Source:HGNC Symbol;Acc:HGNC:41652]	MIR3064	MI0017375	hsa-mir-3064
219811_at	ENSG00000221366	microRNA 1306 [Source:HGNC Symbol;Acc:HGNC:35371]	MIR1306	MI0016008	hsa-mir-3618
229185_at	ENSG00000265304	microRNA 3650 [Source:HGNC Symbol;Acc:HGNC:38981]	MIR3650	MI0016050	hsa-mir-3650
204881_s_at	ENSG00000266315	microRNA 4668 [Source:HGNC Symbol;Acc:HGNC:41545]	MIR4668	MI0017298	hsa-mir-4668
201113_at	ENSG00000264455	microRNA 4721 [Source:HGNC Symbol;Acc:HGNC:41609]	MIR4721	MI0017356	hsa-mir-4721
227488_at	ENSG00000208005	microRNA 503 [Source:HGNC Symbol;Acc:HGNC:32138]	MIR503	MI0003188	hsa-mir-503
1557293_at	ENSG00000207650	microRNA 570 [Source:HGNC Symbol;Acc:HGNC:32826]	MIR570	MI0003577	hsa-mir-570
208922_s_at	ENSG00000274066	microRNA 6514 [Source:HGNC Symbol;Acc:HGNC:50147]	MIR6514	MI0022226	hsa-mir-6514
228089_x_at	ENSG00000274856	microRNA 6748 [Source:HGNC Symbol;Acc:HGNC:50141]	MIR6748	MI0022593	hsa-mir-6748
201782_s_at	ENSG00000276769	microRNA 6752 [Source:HGNC Symbol;Acc:HGNC:50020]	MIR6752	MI0022597	hsa-mir-6752
209572_s_at	ENSG00000273630	microRNA 6755 [Source:HGNC Symbol;Acc:HGNC:50224]	MIR6755	MI0022600	hsa-mir-6755
224739_at	ENSG00000276753	microRNA 6821 [Source:HGNC Symbol;Acc:HGNC:49980]	MIR6821	MI0022666	hsa-mir-6821
58696_at	ENSG00000276472	microRNA 6847 [Source:HGNC Symbol;Acc:HGNC:50022]	MIR6847	MI0022693	hsa-mir-6847
200936_at	ENSG00000274673	microRNA 6850 [Source:HGNC Symbol;Acc:HGNC:50093]	MIR6850	MI0022696	hsa-mir-6850
200775_s_at	ENSG00000207603	microRNA 7-1 [Source:HGNC Symbol;Acc:HGNC:31638]	MIR7-1	MI0000263	hsa-mir-7-1
242246_x_at	ENSG00000211574	microRNA 770 [Source:HGNC Symbol;Acc:HGNC:33143]	MIR770	MI0005118	hsa-mir-770

Probe is the respective probe ids resulted from EC. **Gene-id** is the ensembl id of the genes. **Description** gives the details about the genes. **Gene-name** is the respective gene names. **Gene-source** is the external gene source for each gene. **mirbase-id** is the respective microRNA id for each probe-id.

Table 5.11: List of dysregulated pathways related with SFG region

Pathway Name	**EASE score**	**Gene List**
Nucleotide Metabolism	3.41E-03	DCTD; PRPS1
Metabolism of Cofactors and Vitamins	3.60E-03	NMNAT2; PTPRM
Carbohydrate Metabolism	4.23E-03	PDHA1; PRPS1
Amino Acid Metabolism	5.55E-03	IVD; PDHA1
Phosphatidylinositol signaling system	1.23E-02	PTPRM
Valine, leucine and isoleucine degradation	1.35E-02	IVD
Purine metabolism	1.48E-02	PRPS1
Nicotinate and nicotinamide metabolism	1.60E-02	NMNAT2
Signal Transduction	1.72E-02	PTPRM
Pyrimidine metabolism	1.84E-02	DCTD
Cell Growth and Death	1.96E-02	PTPRM
Butanoate metabolism	2.08E-02	PDHA1
Apoptosis	2.20E-02	PTPRM
Sorting and Degradation	2.32E-02	DCTD
Glycolysis / Gluconeogenesis	2.45E-02	PDHA1
Valine, leucine and isoleucine biosynthesis	2.57E-02	PDHA1
Pyruvate metabolism	2.69E-02	PDHA1
Cell cycle	2.81E-02	PTPRM
Pentose phosphate pathway	2.93E-02	PRPS1
Type IIsecretion system	3.05E-02	DCTD
Riboflavin metabolism	3.17E-02	PTPRM
Biodegradation of Xenobiotics	3.29E-02	PTPRM
gamma-Hexachlorocyclohexane degradation	3.42E-02	PTPRM

Name is the name of the pathways. **EASE Score** is the pathway score by annotating using EASE. **Gene List** is the genes that belongs to each pathway.

Table 5.12: List of microRNAs resulted from SFG region.

Probe-id	Gene-id	Description	Gene-name	Gene-source	mirbase-id
212674_s_at	ENSG00000221585	microRNA 1226 [Source:HGNC Symbol;Acc:HGNC:33922]	MIR1226	MI0006313	hsa-mir-1226
224598_at	ENSG00000221394	microRNA 1229 [Source:HGNC Symbol;Acc:HGNC:33924]	MIR1229	MI0006319	hsa-mir-1229
226144_at	ENSG00000223244	microRNA 1909 [Source:HGNC Symbol;Acc:HGNC:35393]	MIR1909	MI0008330	hsa-mir-1909
220451_s_at	ENSG00000266463	microRNA 3196 [Source:HGNC Symbol;Acc:HGNC:38198]	MIR3196	MI0014241	hsa-mir-3196
208750_s_at	ENSG00000264944	microRNA 3620 [Source:HGNC Symbol;Acc:HGNC:38917]	MIR3620	MI0016011	hsa-mir-3620
225375_at	ENSG00000264302	microRNA 4723 [Source:HGNC Symbol;Acc:HGNC:41660]	MIR4723	MI0017359	hsa-mir-4723
210778_s_at	ENSG00000265080	microRNA 4800 [Source:HGNC Symbol;Acc:HGNC:41877]	MIR4800	MI0017448	hsa-mir-4800
1294_at	ENSG00000263506	microRNA 5193 [Source:HGNC Symbol;Acc:HGNC:43534]	MIR5193	MI0018172	hsa-mir-5193
213156_at	ENSG00000207770	microRNA 568 [Source:HGNC Symbol;Acc:HGNC:32824]	MIR568	MI0003574	hsa-mir-568
1557293_at	ENSG00000207650	microRNA 570 [Source:HGNC Symbol;Acc:HGNC:32826]	MIR570	MI0003577	hsa-mir-570
209111_at	ENSG00000277264	microRNA 6833 [Source:HGNC Symbol;Acc:HGNC:50245]	MIR6833	MI0022678	hsa-mir-6833
58696_at	ENSG00000276472	microRNA 6847 [Source:HGNC Symbol;Acc:HGNC:50022]	MIR6847	MI0022693	hsa-mir-6847
200775_s_at	ENSG00000207603	microRNA 7-1 [Source:HGNC Symbol;Acc:HGNC:31638]	MIR7-1	MI0000263	hsa-mir-7-1

Probe is the respective probe ids resulted from EC. **Gene-id** is the ensembl id of the genes. **Description** gives the details about the genes. **Gene-name** is the respective gene names. **Gene-source** is the external gene source for each gene. **mirbase-id** is the respective microRNA id for each probe-id.

The visual cortex (VCX):

VCX is a part of cerebral cortex that occupies the entire surface of occipital lobe and functions as a visual data receiver. Damage of VCX can make the patient effectively blind even if their eyes are sending information from the visual field to VCX [33, 243]. Even though some studies shows changes in VCX related with normal aging, there is almost no information about changes in VCX in relation with AD [32]. In our study, we could identify a comparatively small number of differentially expressed features for VCX when compared with other regions. Even though we found 2185 differentially expressed features, only 11 have a BF-value < 0.0001, mapping to 10 genes and no significant pathways with EASE score < 0.06. Among the resulted list of features, we find 3 microRNAs that are differentially expressed in VCX. The details of microRNAs are given in Table 5.13.

Table 5.13: List of microRNAs resulted from VCX region.

Probe	Gene-id	Description	Gene-name	Gene-source	mirbase-id
224598_at	ENSG00000221394	microRNA 1229 [Source:HGNC Symbol;Acc:HGNC:33924]	MIR1229	MI0006319	hsa-mir-1229
218466_at	ENSG00000263462	microRNA 4750 [Source:HGNC Symbol;Acc:HGNC:41765]	MIR4750	MI0017389	hsa-mir-4750
238850_at	ENSG00000273878	microRNA 9-2 [Source:HGNC Symbol;Acc:HGNC:31642]	MIR9-2	MI0000467	hsa-mir-9-2

Probe is the respective probe ids resulted from EC. **Gene-id** is the ensembl id of the genes. **Description** gives the details about the genes. **Gene-name** is the respective gene names. **Gene-source** is the external gene source for each gene. **mirbase-id** is the respective microRNA id for each probe-id.

Probes shared among individual region signatures:

Studies have shown that VCX is a less metabolically affected region and exhibits the lowest amount of AD-related changes and gene associations [209, 10, 171, 202, 208]. The individual region analysis here also found that VCX shows fewer common genes with other regions. Despite the large size of individual region signatures, there were only 67 common genes among the individual signatures. The VCX region did not show a large overlap with other regions, and removal of this region's signature increased the number of common genes to 288 among the five other region signatures. Based on this, the VCX region was excluded from the combined analysis.

5.3 Integrated data analysis

The Coloured (α, β)-k Feature Set approach was used for the integrated analysis of the selected datasets and compared our results with two other popular meta-analysis methods: RankProd [140] and GeneMeta [221] (I refer the reader to Chapter 2).

5.3.1 Integration method

In the AD case, the datasets were generated using the same platform, Affymetrix. Thus, combining datasets at the probe level is not an issue as the platform used a constant microarray to generate the gene expression values. Here the datasets were combined directly without applying an integration method.

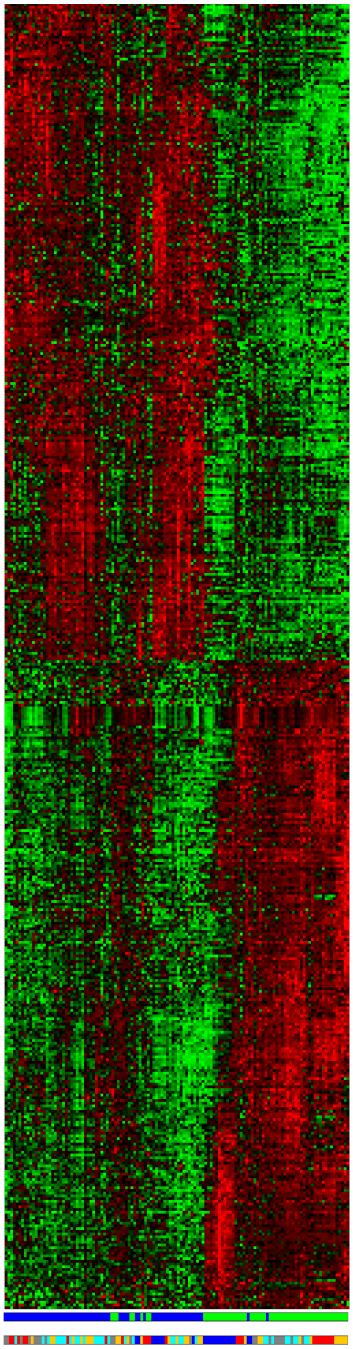
5.3.2 Application of the Coloured (α, β)-k Feature Set problem approach

The proposed combinatorial optimisation-based method, the Coloured (α, β)-k Feature Set problem approach [284] (refer Chapter 3) can handle the combined datasets in a consistent manner and selects the minimum set of significant features that can differentiate sample pairs across multiple datasets.

In the present case, the different datasets correspond to different brain regions. To perform the combined analysis, a combined dataset was prepared by combining all five regions (EC, HIP, MTG, PC and SFG) and selecting the probes that pass Fayyad and Irani's entropy-based heuristic test. The combined dataset containing 3120 features and 126 samples was analysed in two ways, as samples from some individuals (i.e. genotypes) were represented in more than one dataset (brain region). First, the Coloured (α, β)-k Feature Set problem approach (I refer the reader to Subsection 5.3.2 for more details) was applied to the combined dataset, resulting in a list of 825 differentially expressed features with maximum α and β values of 396 and 300, which was annotated to 728 genes. In this list, 479 genes had a BF-value < 0.0001, mapping to 67 significant pathways with an EASE score < 0.06. The list of resulting features and associated details is provided in Table 8.2 and pathways are shown in Table 8.10. The heatmap for 479 features is provided in Figure 5.1.

In the next step, the Generalised (α, β)-k Feature Set problem approach (refer Chapter 3 Section 3.4) was applied to the combined dataset to test whether correlations among samples in different regions might provide a more

Figure 5.1: Heatmap for the 479 features with BF-value < 0.0001 in the brain region-combined analysis. There were 479 up- and down-regulated genes that were differentially expressed between control and AD samples. The first colour bar at the bottom indicates AD (blue) and control (red) samples. The second colour bar represents each sample group in different colours: EC (blue), HIP (red), MTG (orange), PC (grey) and SFG (cyan). The magnified version of the figure is available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4822961/figure/pone.0152342.g001/



robust disease signature. This resulted in a list of 871 differentially expressed features with maximum α and β values of 396 and 311, which was annotated to 747 genes. In this list, the 540 genes had a BF-value < 0.0001, mapping to 70 significant pathways with an EASE score < 0.06. The list of features with associated details and pathways is provided in Table 8.2 and Table 8.11. The heatmap for the 540 features is shown in Figure 5.2. The Coloured (α, β)-k Feature Set and the Generalised (α, β)-k Feature Set problem approach results had 473 genes in common, indicating a high level of agreement between the results. From the increased *beta* value it was deduced that the association of samples across different regions provided a slight increase in the intra-class coherency of the description.

From the lists of features resulting from application of the Coloured (α, β)-k Feature Set and the Generalised (α, β)-k Feature Set problem approaches, 23 non-coding features were identified as being differentially expressed across the EC, HIP, MTG, PC and SFG regions (see Table 5.14). The pathway analysis of these non-coding features identified 13 pathways as shown in Table 5.15. The non-coding features included four microRNAs: hsa-mir-7-1, hsa-mir-570, hsa-mir-1229 and hsa-mir-6821 discussed further in Section 5.6. A heatmap for the 23 features is shown in Figure 5.3.

Figure 5.2: Heatmap for the 540 features with BF-value < 0.0001 from the combined analysis. There were 540 up- and down-regulated probes that were differentially expressed between control and AD sample. The first colour bar at the bottom indicates AD (blue) and control (green) samples. The second colour bar represents each sample group in a different colour: EC (blue), HIP (red), MTG (orange), PC (grey) and SFG (cyan). The magnified version of the figure is available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4822961/figure/pone.0152342.g002/

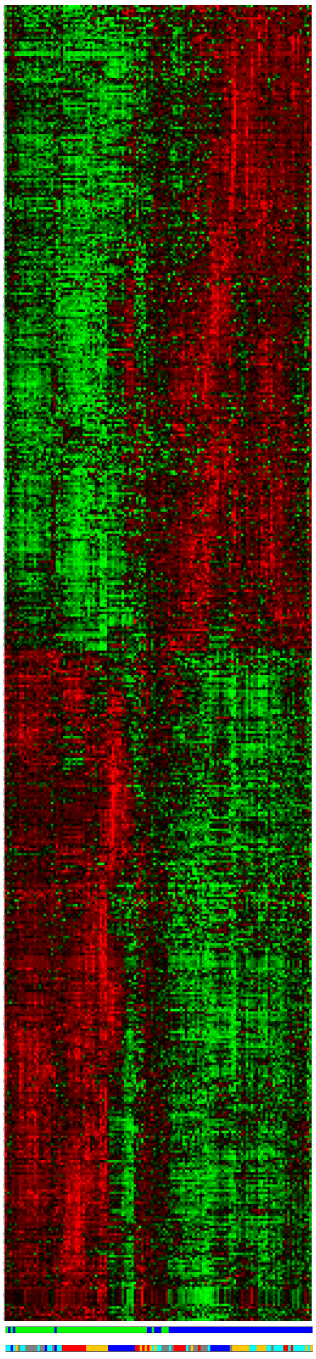


Table 5.14: List of non-coding features resulted from the Coloured (α, β)-k Feature Set and Generalised (α, β)-k Feature Set problem approach.

Probe-id	Target-id	Target-name	Target-source
1557293_at	ENST00000384917; ENST00000445707	MIR570; LINC00969	MIR570-201;LINC00969-001
1558678_s_at	ENST00000534336; ENST00000508832; ENST00000618925	MALAT1	MALAT1-001; MALAT1-002; MALAT1-009
1560982_at	ENST00000569291	RP11-452L6.1	RP11-452L6.1-001
1561346_at	ENST00000579972	RP11-433M22.2	RP11-433M22.2-001
1568763_s_at	ENST00000570130	RP11-480A16.1	RP11-480A16.1-001
200775_s_at	ENST00000384871	MIR7-1	MIR7-1-201
208687_x_at	ENST00000364009	SNORD14E	SNORD14E-201
210679_x_at	ENST00000538710; ENST00000616576	RP11-87C12.5	RP11-87C12.5-001; RP11-87C12.5-002
212384_at	ENST00000617927	AL662801.1	AL662801.1-201
215514_at	ENST00000609675	RP4-621B10.8	RP4-621B10.8-001
223940_x_at	ENST00000534336; ENST00000508832; ENST00000616527	MALAT1	MALAT1-001; MALAT1-002; MALAT1-012
224187_x_at	ENST00000364009; ENST00000534336; ENST00000619449; ENST00000544868	SNORD14E; MALAT1	SNORD14E-201; MALAT1-001; MALAT1-004; MALAT1-007
224567_x_at	ENST00000534336; ENST00000619449; ENST00000544868	MALAT1	MALAT1-001; MALAT1-004; MALAT1-007
224568_x_at	ENST00000534336; ENST00000544868; ENST00000610481; ENST00000508832	MALAT1	MALAT1-001; MALAT1-007; MALAT1-008; MALAT1-002
224598_at	ENST00000408467	MIR1229	MIR1229-201
224739_at	ENST00000617625	MIR6821	MIR6821-201
225055_at	ENST00000620266	RP11-147L13.11	RP11-147L13.11-001
225239_at	ENST00000501122	NEAT1	NEAT1-001
228839_s_at	ENST00000605920; ENST00000609350	RP11-182L21.6; RP11-395A13.2	RP11-182L21.6-001; RP11-395A13.2-001
234989_at	ENST00000501122	NEAT1	NEAT1-001
234997_x_at	ENST00000566446	RP11-488L18.10	RP11-488L18.10-001
239629_at	ENST00000459460	RNU7-45P	RNU7-45P-201
35436_at	ENST00000408370	AL590708.1	AL590708.1-201

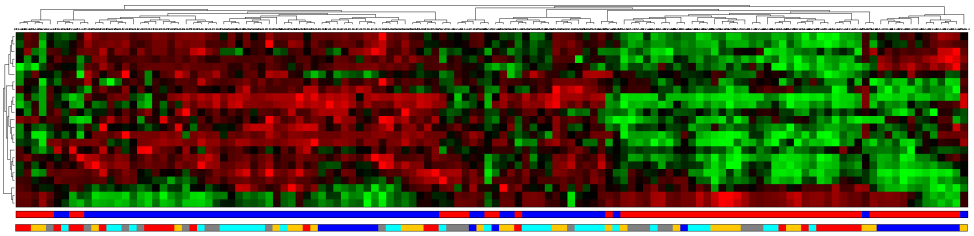
Probe is the respective probe ids resulted from the analysis. **Target-id** is the ensembl id of the target of each microRNA. **Target-name** is the respective target names. **Target-source** is the external source for each target.

Table 5.15: Dysregulated pathway related to non-coding RNAs.

Probe ID	**Symbol**	**Name**	**Pathway**
239629_at	RNU7-45P	RNA, U7 small nuclear 45 pseudogene	Apoptosis, FAS signalling pathway
208687_x_at, 224187_x_at	SNORD14E; MALAT1	small nucleolar RNA, C/D box 14E; metastasis associated lung adenocarcinoma transcript 1	Spliceosome, MAPK signaling pathway, Endocytosis, Antigen processing and presentation, parkinson disease, Membrane Trafficking
200775_s_at	MIR7-1	microRNA 7-1	Spliceosome, Processing of Capped Intron-Containing Pre-mRNA, Influenza Infection, Gene Expression,
224598_at	MIR1229	microRNA 1229	N-Glycan biosynthesis,
1560982_at	CNKSR3	CNKSR Family Member 3	Tight junction, Signalling by GPCR

Probe ID is the probe id related to the respective non-coding features. **Symbol and Name** is the associated target in relation to the non coding RNA that is involved in the pathway. **Pathway** is the name of the pathway.

Figure 5.3: Heatmap for the 23 non-coding features resulting from the combined analysis. These 23 up- and down-regulated features were differentially expressed between control and AD samples. The first colour bar at the bottom indicates AD (blue) and control (red) samples. The second colour bar represents each sample group in a different colour: EC (blue), HIP (red), MTG (orange), PC (grey) and SFG (cyan). The magnified version of the figure is available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4822961/figure/pone.0152342.g003/



A comparison between the Coloured (α, β)-k Feature Set problem approach list and that of brain-related genes in the Human Protein Atlas indicated that 328 genes identified in the current study are normally expressed in human brain but dysregulated in AD. An additional 17 of these genes are not expressed in normal human brain but appear to be over expressed in AD patients.

5.4 Comparison

As comparison benchmark we have used two other widely popular meta-analysis methods RankProd [140] and GeneMeta [41] (I refer the reader to Chapter 2.

The data from the five regions were combined without any pre-processing, as the data were obtained using a single platform. RankProd was applied to the combined region data to select genes that can discriminate between control and AD samples.

The application of RankProd to the combined dataset resulted in a list of genes ordered by increasing percentage of false positive likelihood value, from which the top up- and down-regulated genes (using a 0.05 cut-off) were selected. This resulted in a list of 6908 up-regulated genes and 5853 down-regulated genes. The comparison of the Coloured (α, β)-k Feature Set problem approach result with that of RankProd indicated that 760 of 825 probes were present in the top list of RankProd. Further, the comparison of the generalised Coloured (α, β)-k Feature Set approach result with the RankProd result indicated that 802 of 871 features were present in the results of RankProd. This indicates a high level of agreement between the results achieve using the

different approaches. The features identified by RankProd and the proposed feature selection approaches are shown in Table 8.5.

The data from all five regions were combined without any pre-processing and GeneMeta (please refer Chapter 2 for more details) was applied to identify genes that were differentially expressed in the combined region data.

Application of the GeneMeta method to the combined dataset resulted in the selection of 14,991 features using an FDR cut-off of 0.025. A comparison of the Coloured (α, β)-k Feature Set approach results with those of GeneMeta showed that 684 out of 825 features were present in the GeneMeta list. The comparison involving the Generalised (α, β)-k Feature Set indicated 742 out of 871 features were on the GeneMeta list. Thus, there was a high level of agreement among the results from the three approaches. A comparison of the GeneMeta features with the results from the two (α, β)-k Feature Set approaches is provided in Table 8.13.

5.5 Sensitivity Analysis

The robustness of the final integration results was analysed with respect to varying compositions of the individual region data. This involved repeating the above steps with different combinations of region data prepared by removing single or multiple regions from the combined data with a random selection. This step helps to identify the most significant genes that are not dependent on a single region, as well as each region's contribution to the final results. Specifically, the following steps were performed: a) removal of the EC region from the combined dataset; b) removal of the HIP region from the combined dataset; and (c) removal of the MTG and PC regions from the combined dataset. The resulting list of features that can distinguish classes by applying the Coloured (α, β)-k Feature Set problem approach in each case was compared with the original result.

The results of all three cases indicate an approximate 50% agreement with the original combined region data result. This indicates that the set of resulting genes is statistically significant in the case of the combined dataset and connected with each region. The list of resulted genes after removing different regions from the analysis is provided in Table 8.14. This process pointed to a highly robust result that does not depend on a single dataset.

5.6 Discussion

As AD progresses, tau pathology beings spreading from one brain region to another, in a consistent pattern from the EC to the HIP and then the cerebral cortex. These brain regions are interconnected through synapses that create communication networks [214]. Studies have demonstrated that AD is strongly associated with alterations in connectivity among brain regions [51]. Although studies of gene expression changes associated with AD have been performed separately for different brain regions, a combined study to understand the overlap and difference between different brain regions has been lacking. The current research assessed the differential expression of genes through combined analysis of five brain regions.

The current study has helped identify a set of genes and related pathways that may play an important role in the development of AD. The individual analysis of each of the EC, HIP, MTG, PC, SFG and VCX regions provided sets of genes and pathways that were highly significant for a single region. A set of genes highly associated with all the regions was identified by the combined analysis of EC, HIP, MTG, PC and SFG. As outlined, the VCX region was not included in the combined study because the individual results had little in common with the other regions and thus the inclusion of VCX would have restricted the amount of common evidence that could be obtained from the integrated analysis.

The analysis also identified a small common set of microRNAs that may play a role in AD pathology. The comparison and sensitivity analyses revealed high agreement with the current results. The top-ranked pathways and genes resulting from the combined study are discussed below in detail.

Only two pathways and 67 genes appeared significant when genes common to the results from the region-specific analysis results for the EC, HIP, MTG, PC and SFG regions were annotated. The common pathways were the metabolism of cofactors and vitamins, and the sorting and degradation pathways.

Even though there only two pathways were in common to the regions based on the individual analysis results, the combined analysis using the Coloured (α, β)-k Feature Set problem approach produced 62 dysregulated pathways across five different regions that appeared to be associated with AD. We discuss here the top 15 altered pathways given in Table 5.16.

Table 5.16: Top dysregulated pathways

Pathway	**Gene Symbol**
Carbohydrate Metabolism	ACACB; ALDOA; GRHPR; IDH3A; IDH3B; IDH3G; ME3; PDHA1; PFKFB3; PFKM; PRPS1; TPI1;ABAT; AKR1B1
Valine, leucine and isoleucine biosynthesis	PDHA1
Glyoxylate and dicarboxylate metabolism	GRHPR
MAPK signalling pathway	EGFR
Lipid Metabolism	ACAA1; ACACB; HSD17B7; NQO2
Alzheimer's disease	LPL; SNCA; GNG3
Galactose metabolism	PFKM; AKR1B1
Citrate cycle (TCA cycle)	IDH3A; IDH3B; IDH3G
ATP synthesis	ATP5B; ATP5G1; ATP5J2; ATP6V0B; ATP6V1E1; ATP6V1F
Cell Communication	CAPNS1; SORBS1; TLN2; CAV2
Glutathione metabolism	GSS
Pentose phosphate pathway	ALDOA; PFKM; PRPS1
Amino Acid Metabolism	ABAT; ACAA1; ADSL; EPRS; GFPT1; GOT2; GSS; PDHA1
Arginine and proline metabolism	EPRS; GOT2
Signal Transduction	DGKG; EGFR; INPP4A; ITPR2; PRKCA; PRKCE; PRKCZ; PTPN2;PTPN3; PTPRD; PTPRM

Pathway is the pathway name. **Gene Symbol** are the gene symbols.

The identified pathways are mainly related to the classes of carbohydrate metabolism, amino acid metabolism, signal transduction and lipid metabolism. Carbohydrates are the source of energy that maintains the life of living cells. Carbohydrate metabolic pathways have been previously implicated in AD progression. Henderson [137] showed that consumption of a high-carbohydrate diet may be a cause of the primary event that leads to the development of AD. A series of studies have suggested that relatively simple preventative measures such as lower consumption of starchy carbohydrates and high levels of essential fatty acids in the diet may effectively prevent AD [96, 137, 247, 11, 276, 399, 318]. Moreover, some studies have suggested that carbohydrate diets can lead to dysregulation of lipoprotein lipase (LPL) activity and increase insulin sensitivity [137]. The present study has also suggested under expression of the LPL gene that participates in the AD pathway.

Increased activity of the glycolysis, galactose metabolism [385, 320] and pentose phosphate pathways [272, 31] has also been found to be associated with increased AD risk. All these studies indicate that the risk of AD can be reduced by following a balanced diet of protein, carbohydrate and fat.

Amino acids are the building blocks of proteins, and the dysregulation of amino acid processing can result from defects either in the breakdown of amino acids or in the transport of amino acids into cells. The over representation of amino acid metabolism has also been reported to be associated with AD [116, 229, 113, 211].

The mitogen-activated protein kinase ($MAPK$) signalling pathway has three components: *MAP3K, MAP2K and MAPK*. This pathway regulates a variety of cellular activities including proliferation, differentiation, survival and death. Deviation in the MAPK signalling pathway has been reported in relation to AD [176, 259, 215].

Lipids play a major role in cell signalling, especially in the brain, and are the major energy reserve in the brain cells and tissues. Studies have shown that abnormal lipid metabolism contributes to the pathogenesis of AD and other neurodegenerative disorders [48, 308, 167, 232]. According to the literature search, all the resulting pathways are closely associated with the development of AD.

The discussion now turns to the top-ranked 15 DEGs across all five regions from the results of the Coloured (α, β)-k Feature Set analysis and all the genes are discussed in terms of Coloured (α, β)-k Feature Set problem approach result, shown in Table 5.17.

Table 5.17: The top most DEGs across EC, HIP, MTG, PC and SFG.

Probe ID	**Gene Symbol**	**Gene Name**	**Location**
210976_s_at	PFKM	phosphofructokinase, muscle	12q13.11
200039_s_at	PSMB2	proteasome (prosome, macropain) subunit, beta type, 2	1p34.2
211993_at	WNK1	WNK lysine deficient protein kinase 1	12p13.3
221476_s_at	RPL15	ribosomal protein L15	3p24.1
211921_x_at	PTMA	prothymosin, alpha	2q37.1
46665_at	SEMA4C	sema domain, immunoglobulin domain (Ig), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 4C	2q11.2
223319_at	GPHN	gephyrin	14q23.3
208732_at	RAB2A	RAB2A, member RAS oncogene family	8q12.1
213555_at	RWDD2A	RWD domain containing 2A	6q15
203146_s_at	GABBR1	gamma-aminobutyric acid (GABA) B receptor, 1	6p21.3
224567_x_at	MALAT1	metastasis associated lung adenocarcinoma transcript 1	11q13.1
212296_at	PSMD14	proteasome (prosome, macropain) 26S subunit, non-ATPase, 14	2q14.3
200708_at	GOT2	glutamic-oxaloacetic transaminase 2, mitochondrial	16q21
204786_s_at	IFNAR2	interferon (alpha, beta and omega) receptor 2	21q22.1
215543_s_at	LARGE	like-glycosyltransferase	22q12.3

Probe ID is the probe id that targets the gene. **Gene Symbol** is the symbol of that particular gene. **Gene Name** is the expansion of gene symbol. **Location** is the chromosome location of the gene.

The *PFKM* gene codes for the enzyme phosphofructokinase and catalyses the phosphorylation of fructose-6-phosphate to fructose-1,6-bisphosphate (F-1,6-BP).

F-1,6-BP is broken down into glyceraldehyde 3-phosphate and dihydroxyacetone phosphate under the catalysis of aldolase (*ALDOA or ALDOC*) enzymes. In the current study, *PFKM* was found to be down regulated in all brain regions. Several studies have pointed to a relationship between increased activity of glycolysis and AD [23, 92, 263]. In the current study, *PFKM* was under expressed in all brain regions. Other researchers have also reported that *PFKM* may be associated with the progression of AD in the EC region [53], and differential expression of PFKM has been studied on different rat brain regions in relation to AD [404]. Brooks et al. [34] reported the down regulation of *ALDOA, ALDOC and PFKM* in AD.

The *PSMB2* gene codes for the protein proteasome subunit beta type 2, which is responsible for the degradation of cytosolic and nuclear proteins in the cell. Proteasomes are a major component of eukaryotic cells and cleave peptides in an ATP/ubiquitin-dependent process in a non-lysosomal pathway. Hence, proteasomes play an important role in the ubiquitin-proteasome system, which is an important mechanism in the regulation of cellular cycles, differentiation, transcription, signalling, cell growth and death [101]. Several studies have indicated that aberrations and deregulations of the ubiquitin-proteasome system contribute to the development of neurodegenerative diseases such as AD [39, 359, 266, 296, 197]. These studies present indirect evidence for the role of *PSMB2* in the pathology of AD. Moreover, the differential expression and co-regulation of *PSMB2* with *RPL30* in the HIP region of mouse brain has been reported in relation to AD[365, 271]. To the best of my knowledge, no previous research has indicated a role for *PSMB2* in AD in humans. The current study shows the down regulation of *PSMB2* across all the studied brain regions in association with AD. These results taken together suggest that further studies on *PSMB2* may lead to new insights into AD development.

WNK lysine deficient protein kinase 1 (*WNK1*) encodes for cytoplasmic serine-threonine kinase, which plays a key role in the regulation of blood pressure by controlling the transport of sodium and chloride ions. There is no evidence available for a relationship between *WNK1* and AD progression. However, there is evidence for MAPK-ERK signalling pathway activation by *WNK1* via stimulation of the epidermal growth factor [388, 342, 172]. As mentioned above, MAPK-ERK signalling is highly associated with the pathogenesis of AD [340, 259, 176]. In 2008, Shekarabi et al. [322] reported that

mutations in the nervous system resulting from the over expression of *WNK1* cause a neurodegenerative disorder called hereditary sensory neuropathy type II. More recently, the over expression of *WNK1* was reported to be associated with schizophrenia, a neurodevelopmental disorder [76]. In the current study, *WNK1* was found to be over expressed across all brain regions and it is worth noting that MAPK signalling was one of the pathways that was dysregulated in all five brain regions in this study.

Ribosomal protein L15 (*RPL15*) encodes for the protein 60S ribosomal L15, which plays a key role in RNA binding. Up regulation of *RPL15* has been reported in association with AD in the HIP region of the brain [165]. The current study also found evidence for the up regulation of *RPL15* in the pathogenesis of AD, not only for the HIP region, but also for EC, MTG, PC and SFG. It has also been reported that *RPL15* is closely associated with Parkinson's disease [178] and other brain disorders [279, 288]. The results reported here, along with these studies suggest that *RPL15* may be an important reference gene for AD pathogenesis across different brain regions.

Prothymosin, alpha (*PTMA*) works as a mediator of immune function by conferring resistance to opportunistic infections such as candidiasis and Kaposi's sarcoma. Recent studies have reported the over expression of *PTMA* in AD [77, 185], and its role in transforming growth factor (*TGF*) α-induced apoptosis and estrogen receptor α-induced proliferation [93]. The current study also identified the up regulation of *PTMA* associated with AD in the EC, HIP, MTG, PC and SFG brain regions.

The sema domain, immunoglobulin domain (Ig), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 4C (*SEMA4C*) gene encodes for the semaphorin-4C protein, which is essential for the activation of the *p38 MAPK* gene. The influence of *SEMA4C* on the nervous system is well known and clear: it plays an important role in the development and plasticity of the central nervous system [152, 245, 183, 381, 136]. *SEMA4C* was found to be over expressed in all brain regions in the current study. The expression of *SEMA4C* was originally identified in the nervous system and it is known to be widely expressed in the brains of embryonic and neonatal mice [151]. *p38 MAPK* is emerging as a new AD treatment strategy, and the dysregulation of *p38 MAPK* in AD has been well defined and much studied [409, 176, 259]. According to the best of my knowledge this is the first combined study that has revealed an association between *SEMA4C* and AD in different brain regions.

The *GPHN* gene codes for a neuronal assembly protein called gephyrin, which activates inhibitory neurotransmitter receptors. Both reduced expression of *GPHN* and synaptic dysfunction have been reported in relation to AD

[3], and plaque-like accumulations of gephyrin in AD was identified by Hales et al. [123]. This study has confirmed the down regulation of *GPHN* across all the brain region studied here.

A member of the RAS oncogene family (*RAB2A*) encodes a Rab family protein that is involved in GTP binding and hydrolysis, and participates in the cell cycle. Only a handful of studies have shown that *RAB2A* is associated with neurodegenerative disorders. For example, in 2014, the dysregulation of *RAB2A* in the HIP region was reported by Parra-Damas et al. [275]. The current study has confirmed the down regulation of *RAB2A* in AD samples from the EC, HIP, MTG, PC and SFG regions.

The RWD domain-containing 2A protein encoded by the *RWDD2A* gene is a conserved region of around 110 amino acid residues. It can be found in many ring finger proteins, DEAD-like helicases and WD-repeat-containing proteins, and is mainly involved in protein interaction. A recent age-related study using a mouse model for AD reported over expression of *RWDD2A* in both the HIP and cortex regions [306]. To the best of my knowledge the current study is the first to report the over expression of *RWDD2A* in five different human brain regions. The gamma-aminobutyric acid B receptor-1 (*GABBR1*) gene encodes the main inhibitory neurotransmitter in the human central nervous system. GABBR1 uses ionotropic receptors to produce fast synaptic inhibition, and metabotropic receptors to produce slow and prolonged inhibitory signals. The gene also plays a key role in hippocampal long-term potentiation, slow wave sleep, muscle relaxation and antinociception. *GABBR1* is part of the major histocompatibility complex (MHC) [141] and an association between MHC and AD has been reported in the literature [174]. The expression of *GABBR1* has been widely studied in relation to brain disorders [402, 73, 196, 219].

In 2005, Iwakiri et al. [154] reported that this gene might contribute to AD pathology in the HIP region by altering the balance between the neurotransmitter systems. The under expression of *GABBR1* in AD was noted in the current study for all five brain regions. Detailed study of this gene may contribute new insights to disease progression as the gene is one of the main transmitters in the nervous system.

Proteasome (prosome, macropain) 26s subunit, non-ATPase, 14 (*PSMD14/RPN11*) is a multi-protein complex that plays an important role in the degradation of ubiquitinated intracellular proteins. In the current study, *PSMD14* was found to be under expressed in all the studied brain regions. Down regulation of *PSMD14* has recently been reported in associated with AD [143] and other brain disorders [270, 298, 188].

Glutamic-oxaloacetic transaminase 2, mitochondrial (*GOT2*) is a pyridoxal phosphate-dependent gene that plays an important role in amino acid metabolism. Several studies have reported the down regulation of *GOT2* in AD [24, 346], which is consistent with results from the current study of all five regions. A recent study of RNA transcripts performed by our group also reported that *GOT2* may play a key role in AD pathology [9].

Interferon (alpha, beta and omega) receptor 2 (*IFNAR2*) encodes type I membrane protein, which is involved in the binding and activation of the receptor that stimulates Janus protein kinases such as *STAT1 and STAT2*. The current study identified the over expression of *IFNAR2* in all five brain regions. Several other studies have also reported the up regulation of *INFAR2* in AD pathology [305, 70, 376, 377, 131, 348].

Like-glycosyltransferase (*LARGE*) is one of the largest genes in the human genome and it encodes glycosyltransferase, a protein that participates in glycosylation of alpha-dystroglycan and the synthesis of glycoproteins. Studies have shown that *LARGE* plays a key role in glycosylation [405]. Glycosylation is closely associated with AD and other neurodegenerative disorders [310, 169, 212, 347]. Here, *LARGE* was found to be under expressed in different brain regions. *LARGE* is also known to be closely associated with other brain disorders like neuronal migration disorder, dystroglycanopathies and muscle-eye-brain disease [398, 239, 114].

Fibroblast growth factor (acidic) intracellular binding protein (*FIBP/FGF*) is an intracellular protein that binds selectively to acidic fibroblast growth factor (aFGF). In the current study, *FIBP/FGF* was up regulated across all brain regions. Many other studies have also reported the dysregulation of *FIBP/FGF* in relation to AD [335, 356, 392, 198].

Gamma-aminobutyric acid A receptor, gamma 2 (*GABRG2*) is the major inhibitory neurotransmitter in the mammalian brain, where it acts as a ligand-gated chloride channel. Several studies have demonstrated the down regulation of *GABRG2* in the HIP region of the brain in association with AD and other neurodegenerative disorders [168, 353, 138, 401]. Down regulation of *GABRG2* was also indicated in the current study, not only in the HIP region but also in the EC, MTG, PC and SFG.

This study also identified 23 non-coding features that were differentially expressed across all the brain regions. Although non-coding RNAs are the least understood, their potential for functionality cannot be disregarded. Thus, the importance of non-coding features in AD is worth noting, as these may act as strong future candidates for diagnostic and therapeutic tools in the clinical

treatment of AD. The pathway analysis of these non-coding features resulted in 13 pathways, as shown in Table 5.15.

The expression of non-coding features has previously been linked to several human diseases including cancer, and neurological disorders. Recently, studies of neural differentiation have reported that non-coding features act as additional players in the development of neurological disorders [344, 236, 235, 237]. In the list of 23 non-coding features, *MALAT1, SNORD14E and NEAT1* have previously been reported to be related to neurodegenerative disorders. The current study revealed the differential expression of *MALAT1* and *SNORD14E* and the over expression of *NEAT1* in five different brain regions. Recent studies have reported that heat-stress-related genes like *SNORD14E* are associated with neurodegenerative disorders such as AD, Parkinson's disease and Huntington's disease [331, 182]. The key role of *NEAT1* has been reported in relation to neuronal activity, growth and branching [237, 287, 324]. In addition, the up regulation of *NEAT1* was reported in Huntington's disease [160, 8, 40]. To the best of my knowledge, the current study is the first to report the over expression of *NEAT1* in relation to AD. Another non-coding feature found here to be over expressed is *RP11-488L18.10*, which has been reported as differentially expressed in the astrocyte cells of AD samples [316].

Among these non-coding features, four are microRNAs that are dysregulated in the five brain regions. MicroRNAs play a key role in the development and function of the nervous system, as 70% of known microRNAs are expressed in the brain [79, 410]. These microRNAs are dynamically regulated during brain development, and target different genes and perform different functions in the brain. A study by Sempere et al. [317] reported a group of 17 microRNAs, including hsa-mir-7-1 (miR-7), that play a key role in neuronal differentiation, maturation and/or survival in humans. miR-7 controls epidermal growth factor receptor-related signalling and promotes cell differentiation [207, 289]. The role of miR-7 in modulating α-synuclein levels in the nervous system has also been reported in relation to AD [225] and Parkinson's disease [58]. Several studies have shown the involvement of miR-7 in brain development and disease [164, 58, 86]. Moreover, *EGFR* and the related MAPK signalling pathway are highly ranked in the list resulting here from the pathway analysis. This shows that miR-7 may have an important role in the development of AD, although further studies are needed. hsa-mir-570 (miR-570) has already been reported in relation with brain aging and neurodegeneration [281]. hsa-mir-1229 (miR-1229) has been identified as a suitable biomarker for colon cancers [267]. The role of miR-1229 has not apparently been previously studied or reported in relation to AD.

Differential expression of miR-1229 across five different brain regions is a novel result from this study, as is the differential expression of the relatively unknown microRNA hsa-mir-6821. Even though the mechanism behind the role of miRNAs in disease development remains controversial, the findings here suggest a possible suppression of various cellular functions through the differential expression of this group of miRNAs. Among them, miR-7 and miR-570 have been the subject of intense study in relation to AD.

The comparison with the list of genes collected from the Human Protein Atlas revealed that 328 genes that are normally expressed in human brain were up or down regulated in AD. In the current analysis, 17 genes that are not expressed in normal human brain were over expressed in AD. The three top-ranked genes with good BF-values are examined here.

BEN domain-containing 5 (*BEND5*) plays an important role in the preservation of nervous system integrity by controlling the passage of harmful substances and inflammatory cells into the brain. The Human Protein Atlas records show that *BEND5* is not expressed in normal human brain, and the current study identified its up regulation in AD. Other studies have also reported the up regulation of *BEND5* in relation to AD [273, 145].

Zinc finger protein 415 (*ZNF415*) plays an important role in the gene expression pathway and was shown here to be over expressed in the five different brain regions. Other studies have also revealed differential expression of *ZNF415* in relation to AD [311].

TSPY-like 5 (*TSPYL5*) plays an important role in cell growth and cellular responses to gamma radiation via regulation of the Akt signalling pathway. The current study showed the up regulation of this gene across different brain regions, a phenomenon reported to be associated with AD [205].

Finally, the combined study of AD datasets identified new candidate genes that are consistently differentially expressed across the five different brain regions under study. Further investigations of *PSMB2, WNK1, RPL15, SEMA4C, RWDD2A and LARGE* may provide new insights into the development of AD, and more research on miR-7 and miR570 may contribute more to understanding of the AD pathology. This study has also shown that there are significant differences in the gene expression levels in different brain regions, suggesting there are unique regional activity patterns of AD-affected brain regions and significant differences in the neurodegenerative mechanisms in each region.

Collectively, these results illuminate the potential of these genes to provide insights into AD pathogenesis and the first suggestion that microRNAs may serve as useful biomarkers for AD severity, even though further study is needed.

It is clear that researchers can benefit from these highly AD-associated genes in their search for biomarkers for AD.

5.7 Conclusion

To address region-specific vulnerability with AD pathology and complexity, I performed a comparative study of five different regions (EC, HIP, MTG, PC, SFG) by applying the Coloured (α, β)-k Feature Set and the Generalised (α, β)-k Feature Set problem approaches. The study showed that a meta-analysis methodology with a clear mathematical interpretation guarantees and leads to a largely improved set of markers of AD. A set of six genes and two miRNAs was revealed that warrant further investigation for their high significance in AD-related processes. Although development of drugs directed to treatment of AD is lagging behind, these new findings may provide new insights into disease mechanisms, and thus make a significant contribution in the near future towards understanding, prevention and cure of AD.

Chapter 6

Conclusions and Future work

In this chapter, a conclusion of our study is given in Section 6.1. Future works are presented in Section 6.2. Section 6.3 provide the information about the availability of data and the bioconductor packages that are used in this study. Other publicly available data sources are given in Section 6.4.

6.1 Conclusion

Feature selection from the integrated dataset play a key role in the identification of relevant genes for biological studies. This thesis focused on applying meta-analysis methods for combining individual studies for the detection of biomarkers to understand the underlying biological and biomedical processes. The basic idea behind a meta-analysis is that there is a common research question behind all similar scientific studies, but which has been measured using different methods. In this thesis, a novel feature selection method for meta-analysis, the Coloured (α, β)-k Feature Set problem approach, is presented in Chapter 3. For the application of the proposed methods, we have selected the microarray gene expression datasets and the results are presented in Chapter 4 and Chapter 5. The proposed method performed very well by optimising all the objective functions under consideration at the same time, that are, the minimisation of the number of selected features, the maximisation of α and β and total coverage.

In Chapter 4, we integrated and analysed six prostate cancer microarray gene expression datasets. As per the individual analysis result, there were only 6 common genes between the datasets (we refer the reader to Chapter 4 for more details). The number of genes shared by the datasets is a very small number as all the selected studies were handling the same research question on prostate cancer and it does not allow us to do any type of pathway or biological

analysis. The application of the proposed integration method provided us with a dataset with a large number of samples and features. The application of Coloured (α, β)-k Feature Set problem approach on the combined dataset resulted in a genetic signature with 120 genes, already shown that most of them are indeed related to prostate cancer. The number of common genes has then been increased from 6 to 120, showing that the proposed method is capable to select the most significant genes from the integrated data. It should be noticed that the datasets were generated using different platforms and are not directly combinable. In that case, the Coloured (α, β)-k Feature Set problem approach treats the dataset as different set of sample groups. The result of optimal solutions of the Coloured (α, β)-k Feature Set problem instances generated is then compared against the popular meta-analysis tool called RankProd. We also provide a functional and pathway analysis of the resulted set of genes. The results show that the Coloured (α, β)-k Feature Set problem approach is able to uncover genes with significant and biologically relevant functions that other non-integrative methods fail to identify. The meta-analysis method RankProd uses the common genes from each dataset to combine the selected datasets that will result in loosing those genes that are not common in all datasets. The comparison between the results of the Coloured (α, β)-k Feature Set problem approach and the RankProd also have a high level of agreement for those genes (For more details we refer the reader to Chapter 4).

In Chapter 5, for the application of the Coloured (α, β)-k Feature Set problem approach, we used five different brain region datasets affected by AD. The datasets were directly combinable as the platform used to generate the datasets are same. The individual analysis of the datasets was resulted with 67 common genes. The application of the Generalised (α, β)-k Feature Set problem approach resulted with 540 genes that cover all the five datasets. We identified six new candidate genes, *PSMB2, WNK1, RPL15, SEMA4C, RWDD2A and LARGE*, that are consistently differentially expressed across five different brain regions. Further investigation on these genes may provide us new insights to the development of AD. We also showed that there are significant differences in the gene expression levels in different brain regions, suggesting that there are unique regional activity patterns of AD affected brain regions and significant differences in the neurodegenerative mechanisms within each region. Most of the important genes in these region-specific patterns also become part of the Coloured (α, β)-k Feature Set problem result, showing that the method respects substructure that may be present in the data. The comparison of our proposed method with the state of the art meta-analysis methods called RankProd and GeneMeta shows a very impressive result. Result of the Coloured

(α, β)-k Feature Set problem approach evidences a high level of agreement with the top listed genes of the RankProd result, where almost 80% of our signature is included in the RankProd's result.

Also we identified two new microRNA precursors, miR-7 and miR570, more research on these may contribute more to the AD pathology, even though the platform of the datasets(For more details we refer the reader to Chapter 5) is not designed for microRNAs.

As in the case of prostate cancer, the RankProd results are considerably larger in size, hindering interpretation. Additionally, as mentioned before, the RankProd artificially reduces the rank of any gene with missing values (escalating its position to the significant side of the list), which: i) restricts applicability to the genes represented in all platforms, and ii) introduces non-linear rank scaling in the presence of scattered missing values. In contrast, the Coloured (α, β)-k Feature Set problem approach adequately handles any amount of missing values (that is, a gene may not be present in a dataset but still be significant to explain a large number of sample pairs in the other datasets), providing a more reliable result. Although not used in our investigation, the Coloured (α, β)-k Feature Set problem approach allows for weights to be assigned to genes and samples independently, and accounts for an external perceived relative confidence in each experimental condition, if so desired. The results of the application of the Coloured (α, β)-k Feature Set and the Generalised (α, β)-k Feature Set problem approach on the AD and prostate cancer datasets confirm that our proposed methods are capable to select the most significant features from the integrated datasets and highlight the benefits of using combinatorial optimisation approach for data integration and data mining.

To evaluate the robustness of the proposed method with respect to perturbations in the data, I performed a series of experiments. The presence of noise in the gene expression data is difficult to estimate, as this depends on platform-specific factors as well as experimental conditions. I thus analysed the robustness of the final integration results with respect to the varying compositions of individual datasets, for different perturbation models, inspired by the leave-one-out approach. In the case of prostate cancer datasets, on average, the results remained the same for more than 97% of the signature list, and the signature size remained essentially the same (less than 0.5% increase in the worst case). In the case of AD datasets, the results showed more than 50% agreement with the sensitivity analysis result. This points to a highly robust result that does not depend on a (small) set of genes, even if they are on the high coverage set.

Altogether, application of the proposed method resulted in a dataset with a large number of samples and a more informative signature than the other popular meta-analysis methods. The results indicated that the method is capable of providing highly significant signatures, even where the individual datasets prior to integration are small and thus lacking in informational content. The method is generic and does not depend on the inherent properties of gene expression data, allowing it to be potentially applied to any dataset where the notions of features, class-based classification and equality between feature values is meaningful. As the availability of datasets increases, this novel and robust method can be used to combine more datasets to identify significant genes.

6.2 Future Work

The results of Chapters 4 and 5 show that our methodology is capable to provide an important contribution to the area of biomarker identification. Our results suggest that further investigation is needed into some lesser-researched prostate cancer and Alzheimer's disease genetic biomarkers which appear as informative as others better known. Since the proposed approach is conceptually simple, generalizable, and can be applied to a wide range of experimental set-ups, the method and findings may be of interest for the research community in general and in particular for those working in biological data meta-analysis fields and it is needless to say its applicability to other fields outside the biological realm is guaranteed.

Both the Coloured (α, β)-k Feature Set and Generalised (α, β)-k Feature Set problem approach is a very general method, due to its generality, it is not only applicable to biological data but also to any dataset characterized by continuously valued features. As we are interested in biological data analysis, some of the future applications of the Coloured (α, β)-k Feature Set problem approach in the field of bioinformatics are presented below.

The main goal of researches in life science is to identify the components that make up the living system and to study the (dys) functioning of the system by understanding the interactions between these components. The collection of biological data facilitate the identification of the elements of life, but the integration of these data is required to understand the interactions and other biological functions of living system. The integration study can be performed using mathematical and computational methods that helps to describe the relationship between these elements. The -omics study, aiming at a comprehensive

view on the role of these components in different conditions. Data integration analysis using computational methods becomes so popular in recent years in the field of -omics study, particularly in the area of biomarker detection. The integration of available information from different datasets to generate a combined result seems reasonable and promising for the reinterpretation of studies of a same disease, extracting common knowledge from a set of studies performed under different protocols/technology, performing in-silico analysis of data acquired in other related – but not same – studies, combining data from different realms of proteomics, transcriptomics, epigenetic modifications, etc. Multi-omic integration analysis can be used to understand the differentiation and heterogeneity of cells in development of disease. For instance, the integration analysis of RNA-seq and CHIP-seq data helps to understand the mechanisms behind human diseases like cancer and to improve the treatment. Also, the study on tissue specific alternative splicing, an emerging area of "integrative bioinformatics", may lead us to understand the role of other genetic modifications.

Towards that end, we anticipate the integration of data from different domains like mRNA, miRNA, proteomics, methylation, other epigenetics, etc. The integration of miRNA and mRNA expression data can help to determine the associations between the expression of microRNAs and in particular the putative targeting gene sets. In the case of miRNA expression data, we have to find the targeting genes for each microRNA by using the target prediction tools like TargetScan version 6.1 [274](http://www.targetscan.org/), microCosm version 5 (formerly miRBase) [115](http://www.ebi.ac.uk/enright-srv/microcosm/htdocs/targets/v5/) and miRecords [386] (http://c1.accurascience.com/miRecords/). After the prediction of targeting genes, the integration of data can be performed by finding the overlapping genes between the datasets. The application of the Coloured (α, β)-k Feature Set problem approach on the integrated miRNA – mRNA dataset may lead to the identification of genes that are involved in the most relevant miRNA-mRNA regulatory interactions.

We also plan to incorporate other forms of clinical information available in datasets. In the case of prostate cancer we can also incorporate pathology information like the Gleason score of the samples and design different sample groups depending on the score. Then we can integrate the data in a similar manner presented in Chapter 4 and apply the Coloured (α, β)-k Feature Set problem approach. This way we can retrieve significant genes that may provide more information on different stages of prostate cancer.

Furthermore, other high-throughput sequencing data such as the RNA-seq, ChIP-seq, ATAC-seq, methyl-seq, etc. can be combined in similar manners to correlate the changes in gene expression profiles and diseases. For instance, performing the variant detection using next generation sequencing data to identify the genes that are related with disease specific mutations.

6.3 Final Notes

The data analysis have been performed in R environment. The source code used for the data integration and analysis are given in Appendix 9. Since one of the primary result of our proposed method while using microarray gene expression data are lists of probes/genes. For the annotation of the list of resulted genes, we have used Bioconductor packages like hgu95av2, hgu133a and hgu133plus2(http://www.bioconductor.org/packages/release/data/annotation/). The results producing from the Bioconductor packages helps users to navigate and understand the biological interpretation of the results.

The availability of the datasets used in Chapter 4 is given below.

- Singh dataset: Available at the Broad Institute Cancer Program Legacy Publication Resources website as separate files for normal and tumour samples.

  http://www.broadinstitute.org/mpr/publications/projects/Prostate_Cancer/prostate_normal_N01-N31.CEL.tar.gz

  http://www.broadinstitute.org/mpr/publications/projects/Prostate_Cancer/prostate_normal_N32-N62.CEL.tar.gz

  http://www.broadinstitute.org/mpr/publications/projects/Prostate_Cancer/prostate_tumor_T01-T30.CEL.tar.gz

  http://www.broadinstitute.org/mpr/publications/projects/Prostate_Cancer/prostate_tumor_T31-T62.CEL.tar.gz

- Welsh dataset: Available at the Genomics Institute of the Novartis Research Foundation.

  www.gnf.org/cancer/prostate

- Uma dataset: Available at ArrayExpress under accession number E-GEOD-6919.

  http://www.ebi.ac.uk/arrayexpress/experiments/E-GEOD-6919/

- L-2695, L-3044 and L-3289 datasets: Available in Gene Expression Omnibus under accession number GSE3933

  http://www.ncbi.nlm.nih.gov/geo/

The datasets used in Chapter 5 is available at the Gene Expression Omnibus under accession number GSE5281.

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE5281

6.4 Data Sources

Some of the publicly accessible databases that can be used to collect biological datasets are,

1. Gene Expression Omnibus (GEO): http://www.ncbi.nlm.nih.gov/geo/
2. Array Express: https://www.ebi.ac.uk/arrayexpress/
3. 1000 genomes: http://www.1000genomes.org/
4. The Cancer Genome Atlas (TGCA): http://cancergenome.nih.gov/
5. Encyclopedia of DNA Elements (ENCODE): https://genome.ucsc.edu/ENCODE/index.html
6. GenBank: http://www.ncbi.nlm.nih.gov/genbank/
7. National Center for Biotechnology Information (NCBI): http://www.ncbi.nlm.nih.gov/

At all stages git repositories are used to keep track of records and all the datasets and related results, that are not included in this thesis, are deposited in the local repository. The supplementary materials like heatmaps of individual dataset analysis results, sensitivity analysis results, other datasets that are used at the beginning stage of my research, the related articles, R source codes used for the analysis, etc are available in the repository.

In the case of prostate cancer study, the list of genes resulted from the individual dataset analysis and the heatmaps for each dataset are not included in this thesis, but available as supplementary materials in the published paper in PLOS One[284]. Also, the individual analysis results and the related documents for the AD related study will be available as supplementary materials in the paper accepted by PLOS One.

Chapter 7

Appendix A

7.1 Glossary

Table 7.1: List of notations.

Notation	Description
m	Number of samples
n	Number of features
s	Single feature
(i, j)	Sample pair
M	Discrete matrix of a dataset
y	Tuple, class label of the sample
p	Number of datasets
c	Set of class labels
$\mathcal{F}$	Set of feature nodes
$\mathcal{A}$	Set of alpha nodes
$\mathcal{B}$	Set of beta nodes
v_s	Feature node
$v_{(i,j)}$	α or β node
$x_{(j,s)}$	Expression value of feature s in sample j
e_{sij}	An edge from feature s to the sample pair (i, j)
$L(i)$, $L(j)$	The class labels of samples i and j
c_i, c_j	The colour of that dataset in which the samples i and j belongs to.
t_p	Test effect of dataset p
r_p	p-value of the the study/dataset p
w_p	Weight of study/dataset p
$V * Z$	Standard normal distribution under the null hypothesis
μ	Common effect size
$\mathcal{E}_p$	Within study error
v_p	Within study variance of study/dataset p

Chapter 8

Appendix B

8.1 List of results

This appendix contains the list of results for all data presented in the previous chapters. Also includes the numerical values for those that were either presented graphically or summarized with average values, to ease the interpretation of the text.

8.1.1 Chapter 2 - Meta-analysis methods

Types of meta-analysis

Meta-analysis studies can be classified in two types. The first type tests the statistical significance of the combined results and the other combines the effect size of the individual studies.

Combining the individual study results

The first test of statistical significance on the combined result is proposed in 1931 by L.H.C.Tippett [354]. Soon after R. A. Fisher proposed a method in 1932 (now called *Fisher's method*) for combining p-values across studies by taking the product of those values[81, 254]. Fisher's test can be based on one-tailed or two-tailed test statistics. Fisher's test works best based on one-tailed test statistics. At the same time, Fisher's test does not favour alternatives with common sign in two-tailed based statistics. Then in 1933, Karl Pearson [278] derived another method for meta-analysis called *Pearson's method* that favour the two-tailed test statistics by combining p-values across studies. There after, meta-analysis by combining individual study results has flourished. Some of the popular meta-analysis methods that are performed by

combining the individual study results are given in the next subsections. First we set a null hypothesis, that can be applied on all the methods explained in this section.

Suppose there are P independent experiments and θ_p are the unknown parameters that characterise the effect of each study p, $p = \{1, ..., P\}$. The null hypothesis of pth experiment is $H_{0p} : \theta_p = 0$. If the overall test T_p has a continuous distribution, the significance of the test can be defined as $S_p = Pr(T_p > t_p|H_{0p})$, called p-value, in which t_p is the test effect of a single study. S_p is uniformly distributed only when H_{0p} is true. The test of the combined statistical significance using p-values is a type of non-parametric meta-analysis test as the p-value does not depend on the distribution of the data. This test depends only on the fact that the p-values are distributed between 0 and 1 under the null hypothesis.

Tippett's statistic

The p-value statistics proposed by Tippett [354] in 1931 is given by,

$$V^{min\,R} = min\, r_p,\ p = \{1, ..., P\}$$

In which r_p is the p-value of the individual study p and is uniformly distributed on the interval [0,1] under the null hypothesis. So the distribution of $V^{min\,R}$ can be easily derived as a *beta* distribution which is a family of continuous probability distributions defined on the interval

$$0, 1$$

and are parametrised by two positive integers,denoted by α and β, that appear as exponents of the random variable and control the shape of the distribution under the null hypothesis with the parameters $\alpha = 1$ and $\beta = P$. That is, H_0 is rejected if $V^{min\,R} < 1 - (1 - \alpha)^{1/p}$ where α is the overall significance level. The maximum p-value statistics is given by,

$$V^{max\,R} = max\, r_p,\ p = \{1, ..., N\}$$

In the same way the distribution of $V^{max\,R}$ can be derived as an *alpha* distribution with the parameters $\alpha = P$ and $\beta = 1$.

Fisher's statistic

This is a well known statistics proposed by Fisher and Mosteller [81, 254], and is given by the expression,

$$V^{Fisher} = -2\sum_{p=1}^{P} log(r_p)$$

Under the null hypothesis, r_p is uniformly distributed to the interval [0,1]. But the distribution of $-log(r_p)$ is a *gamma* distribution which is a two-parameter (shape parameter and scale parameter) family of continuous probability distributions with the parameters $\alpha = 1$ and $\beta = 1$. That makes the distribution of V_{Fisher} a *gamma* distribution with parameters $\alpha = p$ and $\beta = 1/2$.

Weighted Fisher's statistic

The weighted Fisher's statistic is an extended version of Fisher's statistic proposed by Good [106] and is given by the formula,

$$V^{WF} = -\sum_{p=1}^{P} w_p \, log(r_p)$$

in which, w_p is the constant weight for the pth study. The weight of the study is determined using the available information about the study. According to Good's study, the distribution function can be defined as,

$$Pr(V^{WF} < x) = 1 - \sum_{p=1}^{P} \Lambda_p e^{-x/2w_p}$$

where

$$\Lambda_p = \frac{w_p^{N-1}}{\prod_{j=1}^{N}(w_p - w_j)}$$

in which, w_j is the weight of any other study other than p and x is a random variable drawn from an uniform independent distribution.

In 1978, Koziol and Perlman [189] proved that the weighted Fisher's statistic is more powerful than standard Fisher's procedure, even though the exact distribution of weighted Fisher's statistic results in an ill-conditioned calculation whenever any of the weights is zero or any two weights are equal. Li and Tseng [203] proposed a another method considering these issues called adaptively weighted Fisher's statistic.

Adaptively weighted Fisher's statistic

The adaptively weighted Fisher's statistic is given by the formula,

$$V^{AW} = \min_{w_s \in W} r(u_s(w))$$

where

$$u_s(w_s) = -\sum_{p=1}^{P} w_{ps}\, log(r_{ps})$$

in which w_p is the weight for the pth study and $w_p s = (w_{1s}, ..., w_{Ps})$ and s is a feature/gene.

The adaptively weighted Fisher's statistic assigns weights to each individual study depends on the distribution of each data and searches through all possible weights to find the best adaptive weight with the smallest derived p-value. One significant advantage of this method is its ability to indicate which studies contribute to the evidence aggregation and elucidates heterogeneity in the meta-analysis.

Inverse normal statistic

Inverse normal statistic is proposed by Stouffer [336] in 1949, is a method widely used in meta-analysis studies to combine p - values across studies.

Inverse normal statistic is given by the formula,

$$V^{Z} = \frac{\sum_{p=1}^{P} \phi^{-1}(r_p)}{\sqrt{P}}$$

V^Z follows a standard normal distribution under null hypothesis and H_0 is rejected when V^Z is greater than the critical value of the standard normal distribution. The weighted version of the Inverse normal statistic [403] can be derived as,

$$V^{WZ} = \frac{\sum_{p=1}^{P} w_p \phi^{-1}(r_p)}{\sqrt{P}}$$

Combining effect sizes

The meta-analysis methods by combining effect sizes are very useful when studies have comparable designs and measure the outcomes in a similar manner. Effect sizes are quantitative measures that are used to summarise the results of a study used for meta-analysis. That is, effect sizes are the reflection

of the association of the magnitudes of interest in each study. There are many different types of effect sizes used in a meta-analysis that represents the results of a study in an easily interpretable and comparable way. Fixed and random (mixed) effects are the two major types of statistical models that have been developed for inference about effect size data from a collection of studies. The procedures for the data analysis are similar in these two types of meta-analysis but the statistical test are somewhat different.

In fixed effects models, the population values of the treatment effects are fixed. That means there is a true effect size that is shared by all the selected studies for meta-analysis. The simplest fixed effect model involves the estimation of the average effect size which is performed by combining the effect size of individual studies selected for the meta-analysis.

In fixed effect model we assume that the studies included for the analysis share a common effect size μ. The observed effect will be distributed on μ with a variance of σ^2 that depends on the sample size of each study. That is, the observed side effect of a study T_p is given by,

$$T_p = \mu + \mathcal{E}_p$$

In which $\mathcal{E}_p$ is the within study error. We need to handle with only within study error as the approach have only one level of sampling, since all studies are sampled from a population with effect size μ.

Since our goal is to give more importance to the studies that carry more information, we can propose a weight to each study according to its sample size. so that a study with more samples would get more weight in the analysis. This is the approach that basically used, but in some cases the weights are assigned based on the inverse of the variance rather than sample size. The inverse variance is roughly proportional to sample size. In that case the weight of each study will be,

$$W_p = 1/v_p$$

where W_p and v_p are the weight and within study variance of study p. Then the weighted mean of the combined study is the sum of the product of effect size and weight divided by the sum of the weights. That is,

$$T_\cdot = \frac{\sum_{p=1}^{P} W_p T_p}{\sum_{p=1}^{P} W_p}$$

Also the variance of the combined effect is the reciprocal of the sum of the weights,

$$v_\cdot = \frac{1}{\sum_{p=1}^{P} W_p}$$

then the standard error of the combined effect is given by,

$$SE(T_\cdot) = \sqrt{v_\cdot}$$

Finally a Z score for each feature in each study is computed by,

$$Z = \frac{T_\cdot}{SE(T_\cdot)}$$

in which Z-score is the statistical measurement of a score's relationship to the mean in a group of scores. When the mean of the sore is same as the individual score the Z-score will be 0. Z-score can also be positive or negative, indicating that whether it is above or below the mean in terms of standard deviations.

This is the common procedure that use in fixed effect models. There is a wide range of studies available in literature on fixed effect models. Some of them are given below. A detailed review of the fixed effect model is presented by Hedges and Olkin [132] in 1985. Again Hedges [133] and Draper [60] have provided a more updated review on these methods. There are several articles in the literature that use these methods of fixed-effects models. Smith et al. [328] summarises the fixed effect models and presented a comparison study with Mantel-Haenszel [228], Woolf's method [380], Mantel-Haenszel-Peto method [88] and logistic regression. The Mantel-Haenszel method weighted averages of the maximum likelihood estimates of the log(oddsratios) in each study and Woolf's method use the same parameter but the odds ratio instead of log(oddsratio).

In random or mixed effect model, the effect sizes are treated as if they were a random sample from a population of effect parameters [134]. In this case, an unavoidable variation is present among the effect size parameters even after controlling the factors of interest in the study that makes the residual variation from the sampling error greater than expected. The mixed effects model is a generalization of the fixed effects model that incorporates the component of between-study variation into the effect size parameters and their estimates. In which the observed effect of a study is determined by the sum of the true effect of that study and within study error. In this case a true effect is used instead of

assumed effect. Also the simplest random effect model involves the estimation of the average effect size by combining the effect size of each study selected for the meta-analysis. But in random effect model, a two level model is used, one for within-study variation and the other for between-study variation. The within-study level is calculated as the same as fixed model effect. At the same time, the between-study level is calculated using a mean effect size and a study specific random effect. The true effect for a study is calculated by the mean of all true effect and another variable called between study error. That is,

$$T_p = \theta_p + \mathcal{E}_p$$

where θ_p is the true effect and is calculated by,

$$\theta_p = \mu + \delta_p$$

where δ_p is the between study error. Then the random effect model is performed by following the steps given in the fixed effect model.

8.1.2 Chapter 3 - Results

Input file format for Coloured (α, β)-k Feature Set problem and Generalised (α, β)-k Feature Set problem approach.

This file has a fairly fixed format, with keywords and values in list or matrix format. An example is given in Figure 8.1, lines 1 to 5 and the data section are mandatory, and they define the minimum contents of the file. Following lines are optional and they define additional parameters for the problem. Each line consists of a keyword or tokens separated by white space (blanks, tabs). Names can not contain embedded white space, it is recommended to use unique identifiers for feature and case names. Each keyword is discussed according to the line number in the file where the keyword appears and the alternative values for the keyword in parenthesis.

1. FEATURESINROWS (FEATURESINCOLUMNS): This keyword specifies the way the main data is organised (lines 6 to 10): one line per feature, or one line per sample.
2. TARGETPRESENT (TARGETNOTPRESENT): Specifies if sample classes are provided in the file or not.
3. LAST (FIRST, NO): If sample classes are provided and data format is FEATURESINROWS, this keyword specifies the place where the sample classes are provided in the main data: if the last line after the data (line 11 in the example) or the first line after the sample names (line 6 in the example). If the data format is FEATURESINCOLUMNS, the classes are specified by a sub-keyword in the sample names line. If no classes are provided, the keyword NO should be used.

4–5 Number of rows and columns in the main data part. These actually correspond to the number of features and samples (or vice-versa, depending on data format).

Figure 8.1: Example for the input file of Coloured (α, β)-k Feature Set problem and Generalised (α, β)-k Feature Set problem approach

```
FEATURES:NROWS
TARGET:PRESENT
LAST
2100
48
dummy Case1 Case2 ... Case48
F1       0 2 ... 1
F2       0 2 ... 1
... ...
F2100 0 2 ... 1
      A A ... B
FEATUREWEIGHTS
1 2 ... 1
CASEWEIGHTS
10 1 5 ... 1
CASECOLOURS
a a ... d d
ALFACOLOURS
{ true | false }
BETACOLOURS
{ true | false }
FEATURES:N
F1 F3 ...
FEATURESOUT
F5 F14 ...
BETA
0 1
CASEADJACENCY
0 0 1 1 0 ... 0
0 0 1 0 1 ... 0
1 1 0     ... 0
      ...
0 0   ...   1 0
ALFAADJACENCY
{ true | false }
BETAADJACENCY
{ true | false | negate }
```

6 Sample / Feature names. Notice that the first token in this line is a dummy place holder, as the first column has the (features or samples) names. If the special sub-keyword classes is given as the second or last token of the line, this would indicate if the second (or last) column for a FEATURESINROWS formatted data file contains the classes.

7–10 Data. First column contains the feature or sample name, following tokens specify the feature values for each case. Non existing values can be specified with the MISSING-VALUE attribute, currently this is -1. Values should be given as integers or character strings.

11 This line contains the sample classes. Depending on the LAST / FIRST keyword, it could be located after line 6.

12–13 FEATUREWEIGHTS (optional): Indicates the next line provides the weights of each feature (in the same order as the data section). If not given, weights are implicitly assumed to be equal to 1. Weights may be given as floating point numbers.

14–15 CASEWEIGHTS (optional): Indicates next line contains weights for each case (in the same order as in the data section). If not given, weights are implicitly assumed to be equal to 1. Weights may be given as floating point numbers.

16–17 CASECOLOURS (optional): Indicates next line contains the "colours" of the data-sets, as discussed previously in chapter 3. Colours are character strings or numbers in the case of Coloured (α, β)-k Feature Set problem approach. The use of CASE-COLOURS is incompatible with the use of CASEADJACENCY in the same file in the case of Generalised (α, β)-k Feature Set problem approach.If not given, they are implicitly assumed to be equal.

18–21 ALFACOLOURS, BETACOLOURS (optional): Indicates additional restrictions on the creation of alpha, beta pairs. Currently, the values true (default if omitted) and false are supported, indicating if sample pair colour must be the same or different, respectively.

22–23 FEATURESIN (optional): Indicates next line contains the feature names that must be part of the solution.

24–25 FEATURESOUT (optional): Indicates next line contains the feature names that must not be part of the solution.

26–27 BETA (optional): Indicates next line contains the classes to be considered as valid beta nodes, actually restricting the number and type of possible beta nodes in the problem. If not present, then beta nodes for all classes given are constructed.

28–33 CASEADJACENCY (optional): This keyword, if provided, marks the beginning of a symmetric (number of columns) $\times$(number of columns) boolean matrix. Diagonal elements are ignored. The use of CASEADJACENCY is incompatible with the use of CASECOLOURS in the same file.

34–35 ALFAADJACENCY (optional): A convenient keyword to turn "on" (true, the default) or "off" (false) the use of the adjacency matrix for α nodes.

36–37 BETAADJACENCY (optional): A convenient keyword to turn "on" (true, the default) or "off" (false) the use of the adjacency matrix for β nodes.

8.1.3 Chapter 4 - Results

The result of Coloured (α, β)-k Feature Set problem with the details of run time, coverage, etc.

Table 8.1: Details of Coloured (α, β)-k Feature Set problem result.

Features: 16157, Cases:319, Alfa Nodes:4797, Beta Nodes: 5908
Maximum Alfa: 11271, Maximum Beta: 11375, Minimum Alfa: 612, Minimum Beta: 1784
NA value: -1, Classes: 2, Colours: 6
Beta targets: 2
Has Targets: false
Has Case Weights: false
Has Case Colours: true
Has Feature Weights: true
Presolve time = 14.47 sec. (7617.89 ticks)
Probing time = 0.22 sec. (90.62 ticks)
Tried aggregator 1 time.
Presolve time = 5.40 sec. (2608.36 ticks)
Probing time = 0.22 sec. (90.62 ticks)
Parallel mode: deterministic using up to 32 threads.
Root relaxation solution time = 34.38 sec. (23295.45 ticks)
Cover cuts applied: 4
Zero-half cuts applied: 2
Gomory fractional cuts applied: 4
Root node processing (before b&c):
Real time = 446.49 sec. (243124.71 ticks)
Parallel b&c 32 threads:
Real time = 1660.65 sec. (632241.74 ticks)
Sync time (average) = 0.36 sec.
Wait time (average) = 0.00 sec.
Total (root+branch&cut) = 2107.14 sec. (875366.46 ticks)
Solution for model: Max Sum alfa beta cover
Solution status = Optimal
Solution value = 1.55363e+07
Solution time = 26974.8
Condition: Alfa: 612 Beta: 776 K: 3190 Option: gap 0
**Solution: Optimal Cover: 15536314
————————————

The list of Coloured (α, β)-k Feature Set problem approach result that cover six datasets.

Table 8.2: **The list of Coloured (α, β)-k Feature Set problem approach result that cover six datasets.**

Combination of probes	Symbol
L2695:f35369;L3044:f35309;L3289:f35309;Welsh:32634_s_at;Uma:32634_s_at; Singh:32634_s_at	ICA1
L2695:f33094;L3044:f5482;L3289:f35138;Welsh:36484_at;Uma:36483_at; Singh:36483_at	GALNT3
L2695:f16333;L3044:f18929;L3289:f18929;Welsh:32001_s_at;Uma:32001_s_at; Singh:32001_s_at	PCSK6
L2695:f35455;L3044:f34287;L3289:f34287;Welsh:39598_at;Uma:39598_at; Singh:38614_s_at	BCYRN1;GJB1;ZMYM3; NONO;ITGB1BP2;TAF1; INGX;OGT; ACRC; CXCR3;LOC100132741; FLJ46446;CXorf49B;CXo
L2695:f44074;L3044:f16527;L3289:f41754;Welsh:39748_at;Uma:39748_at; Singh:39748_at	SLC7A1
L2695:f31570;L3044:f8994;L3289:f35262;Welsh:39732_at;Uma:39732_at; Singh:39732_at	MAP7
L2695:f32664;L3044:f27309;L3289:f27309;Welsh:33375_at;Uma:33375_at; Singh:33375_at	MYO6
L2695:f2701;L3044:f31476;L3289:f31476;Welsh:39925_at;Uma:39925_at; Singh:39925_at	COL9A2
L2695:f44306;L3044:f7122;L3289:f37581;Welsh:38218_at;Uma:38218_at; Singh:38218_at	GCNT1
L2695:f37438;L3044:f39830;L3289:f39830;Welsh:41706_at;Uma:41706_at; Singh:41706_at	C1QTNF3-AMACR;AMACR; C1QTNF3
L2695:f11412;L3044:f12464;L3289:f12464;Welsh:34050_at;Uma:34050_at; Singh:34050_at	ACSM1
L2695:f14927;L3044:f14831;L3289:f14831;Welsh:36965_at;Uma:36967_g_at; Singh:36965_at	ANK3
L2695:f20946;L3044:f35553;L3289:f35553;Welsh:35214_at;Uma:35214_at; Singh:35214_at	UGDH
L2695:f15006;L3044:f18342;L3289:f13629;Welsh:33131_at;Uma:33131_at; Singh:33131_at	SOX4
L2695:f7291;L3044:f14964;L3289:f2691;Welsh:32034_at;Uma:32034_at; Singh:32034_at	ZNF217
L2695:f41979;L3044:f36211;L3289:f36211;Welsh:39542_at;Uma:39542_at; Singh:39542_at	ENC1
L2695:f11222;L3044:f13054;L3289:f8642;Welsh:36918_at;Uma:36918_at; Singh:36918_at	GUCY1A3
L2695:f40766;L3044:f26446;L3289:f26446;Welsh:33358_at;Uma:33358_at; Singh:33358_at	PPM1H
L2695:f3928;L3044:f7792;L3289:f7792;Welsh:37366_at;Uma:37366_at; Singh:40059_r_at	PDLIM5
L2695:f40064;L3044:f16102;L3289:f16102;Welsh:41468_at;Uma:41468_at; Singh:41468_at	TARP
L2695:f28778;L3044:f39224;L3289:f39224;Welsh:36160_s_at;Uma:36160_s_at; Singh:36160_s_at	PTPRN2;MIR153-2; LOC100506585;MIR595
L2695:f26255;L3044:f17467;L3289:f40139;Welsh:32684_at;Uma:32684_at; Singh:32684_at	C9orf91
L2695:f30148;L3044:f29068;L3289:f29068;Welsh:38716_at;Uma:38716_at; Singh:38716_at	CAMKK2
L2695:f26968;L3044:f25344;L3289:f25344;Welsh:37639_at;Uma:37639_at; Singh:37639_at	SCN1B;HPN; LOC100128675
L2695:f21536;L3044:f20788;L3289:f20788;Welsh:38464_at;Uma:38464_at; Singh:38464_at	INO80B;INO80B-WBP1; WBP1;MOGS
L2695:f9491;L3044:f9251;L3289:f9251;Welsh:37955_at;Uma:37955_at; Singh:37955_at	CNPY2;PAN2
L2695:f37167;L3044:f40627;L3289:f40627;Welsh:374_f_at;Uma:41181_r_at; Singh:33689_s_at	DDTL;DDT;GSTT2
L2695:f24;L3044:f4668;L3289:f4668;Welsh:36587_at;Uma:36587_at; Singh:36587_at	EEF2;SNORD37
L2695:f4266;L3044:f130;L3289:f130;Welsh:40435_at;Uma:40435_at; Singh:40435_at	SLC25A6

The list of Coloured (α, β)-k Feature Set problem approach result that cover six datasets.(continued)

Combination of probes	Symbol
L2695:f3894;L3044:f6337;L3289:f3614;Welsh:31545_at;Uma:31545_at; Singh:31545_at	VPS52;RPS18; B3GALT4;WDR46; PFDN6;RGL2; TAPBP;ZBTB22; DAXX
L2695:f37436;L3044:f39828;L3289:f39828;Welsh:31568_at;Uma:31568_at; Singh:31568_at	RPS10-NUDT3;NUDT3; RPS10
L2695:f16672;L3044:f19748;L3289:f19748;Welsh:33820_g_at;Uma:33819_at; Singh:33820_g_at	LDHB
L2695:f25093;L3044:f11703;L3289:f11703;Welsh:33994_g_at;Uma:38251_at; Singh:33994_g_at	MYL6B;MYL6; SMARCC2
L2695:f6668;L3044:f1804;L3289:f1804;Welsh:32314_g_at;Uma:32314_g_at; Singh:32314_g_at	CA9
L2695:f10260;L3044:f9076;L3289:f9076;Welsh:32243_g_at;Uma:32243_g_at; Singh:32243_g_at	CRYAB;HSPB2; HSPB2-C11orf52;C11orf52
L2695:f3215;L3044:f1883;L3289:f1883;Welsh:38057_at;Uma:38057_at; Singh:38059_g_at	DPT
L2695:f17538;L3044:f34268;L3289:f34268;Welsh:1734_at;Uma:1767_s_at; Singh:1767_s_at	TGFB3
L2695:f24131;L3044:f23547;L3289:f23547;Welsh:37326_at;Uma:37326_at; Singh:37326_at	PLP2;PRICKLE3
L2695:f37082;L3044:f32334;L3289:f32334;Welsh:769_s_at;Uma:757_at; Singh:769_s_at	ANXA2
L2695:f35118;L3044:f16348;L3289:f16348;Welsh:32700_at;Uma:32700_at; Singh:32700_at	GBP2
L2695:f27244;L3044:f21940;L3289:f21940;Welsh:41385_at;Uma:41385_at; Singh:41385_at	EPB41L3
L2695:f2267;L3044:f32879;L3289:f32879;Welsh:39852_at;Uma:39852_at; Singh:39852_at	SPG20
L2695:f121;L3044:f27092;L3289:f27092;Welsh:40099_at;Uma:40099_at; Singh:40099_at	ARHGEF2
L2695:f25104;L3044:f33325;L3289:f27980;Welsh:39701_at;Uma:39701_at; Singh:39701_at	ZIM2;PEG3; PEG3-AS1;MIMT1
L2695:f28978;L3044:f28334;L3289:f28334;Welsh:39550_at;Uma:39550_at; Singh:39550_at	GLT25D2
L2695:f13471;L3044:f35970;L3289:f35970;Welsh:35213_at;Uma:35213_at; Singh:35213_at	WBP4
L2695:f8705;L3044:f12277;L3289:f12277;Welsh:837_s_at;Uma:837_s_at; Singh:837_s_at	ME1
L2695:f37131;L3044:f39659;L3289:f39659;Welsh:39243_s_at;Uma:39243_s_at; Singh:37622_r_at	PSIP1
L2695:f38361;L3044:f36964;L3289:f36964;Welsh:1276_g_at;Uma:1276_g_at; Singh:38047_at	RBPMS
L2695:f9313;L3044:f12601;L3289:f12601;Welsh:38098_at;Uma:38098_at; Singh:38098_at	LPIN1
L2695:f12413;L3044:f21595;L3289:f8249;Welsh:38722_at;Uma:38722_at; Singh:38722_at	COL6A1
L2695:f12638;L3044:f18141;L3289:f18141;Welsh:33862_at;Uma:33862_at; Singh:33862_at	PPAP2B
L2695:f19611;L3044:f26937;L3289:f26937;Welsh:32582_at;Uma:32582_at; Singh:774_g_at	KIAA0430;NDE1; MIR484;MYH11
L2695:f25581;L3044:f13632;L3289:f24386;Welsh:40069_at;Uma:40069_at; Singh:40069_at	SVIL;MIR604; MIR938
L2695:f19545;L3044:f18241;L3289:f18241;Welsh:41744_at;Uma:41743_i_at; Singh:41744_at	OPTN
L2695:f3820;L3044:f4712;L3289:f4712;Welsh:32526_at;Uma:41862_at; Singh:32526_at	JAM3;NCAPD3; VPS26B;THYN1; ACAD8
L2695:f3249;L3044:f423;L3289:f423;Welsh:32851_at;Uma:34532_at; Singh:32851_at	CELF2
L2695:f9562;L3044:f31036;L3289:f31036;Welsh:41870_at;Uma:31617_at; Singh:41871_at	PDPN
L2695:f23246;L3044:f38922;L3289:f30147;Welsh:35776_at;Uma:39096_at; Singh:488_at	GART;SON; DON-SON;CRYZL1; ITSN1

The list of Coloured (α, β)-k Feature Set problem approach result that cover six datasets.(continued)

Combination of probes	Symbol
L2695:f37357;L3044:f40761;L3289:f40761;Welsh:39388_at;Uma:32104_i_at; Singh:650_s_at	KIAA0913; LOC100507331; NDST2;CAMK2G
L2695:f33091;L3044:f35135;L3289:f35135;Welsh:35737_at;Uma:35738_at; Singh:35737_at	HMGN4
L2695:f20532;L3044:f19996;L3289:f19996;Welsh:41143_at;Uma:41288_at; Singh:41144_g_at	CALM1
L2695:f17536;L3044:f17088;L3289:f17088;Welsh:39939_at;Uma:39939_at; Singh:39939_at	PSMD10;ATG4A; COL4A6;COL4A5
L2695:f13413;L3044:f10113;L3289:f10113;Welsh:39366_at;Uma:39366_at; Singh:39366_at	PPP1R3C
L2695:f2786;L3044:f6414;L3289:f6414;Welsh:32076_at;Uma:32076_at; Singh:32076_at	RCAN2

Combination of probes is the list of combined probes. **Symbol** is the list of gene symbols corresponding to each combined probes.

The row names for the heatmap given in Figure 4.3

Table 8.3: **The row names for the heatmap given in Figure 4.3.**

Combination of probes	Symbol
L2695:f22317;L3044:f26225;L3289:f26225;Welsh:1363_at;Singh:1363_at	FGFR2
L2695:f38206;L3044:f29243;L3289:f20298;Welsh:36690_at;Uma:36690_at	NR3C1
L2695:f41990;L3044:f37162;L3289:f37162;Welsh:34693_at;Singh:34693_at	LOC100507246; SNORD1C;SNORD1B; SNORD1A;ST6GALNAC
L2695:f36345;L3044:f35333;L3289:f35333;Welsh:38944_at;Uma:1433_g_at	SMAD3
L3044:f26180;L3289:f26180;Welsh:36943_r_at;Uma:36943_r_at;Singh:36943_r_at	PLAGL1
L2695:f37082;L3044:f32334;L3289:f32334;Welsh:769_s_at;Uma:757_at;Singh:769_s_at	ANXA2
L2695:f39538;L3044:f9089;L3289:f24623;Welsh:36591_at;Uma:36591_at	ZFAND2B;ABCB6;ATG9 ANKZF1;GLB1L;STK16; TUBA4A;TUBA4B
L3044:f23449;L3289:f23449;Welsh:38653_at;Uma:38653_at;Singh:38653_at	PMP22
L2695:f9308;L3044:f9904;L3289:f9904;Welsh:38756_at;Uma:1848_at	RAP1A
L2695:f21134;L3044:f19966;L3289:f19966;Welsh:41739_s_at;Uma:41739_s_at	CALD1
L2695:f6668;L3044:f1804;L3289:f1804;Welsh:32314_g_at;Uma:32314_g_at; Singh:32314_g_at	CA9
L2695:f4800;L3044:f4740;L3289:f4740;Welsh:41839_at;Uma:41839_at	GAS1
L2695:f16672;L3044:f19748;L3289:f19748;Welsh:33820_g_at;Uma:33819_at; Singh:33820_g_at	LDHB
L2695:f25093;L3044:f11703;L3289:f11703;Welsh:33994_g_at;Uma:38251_at; Singh:33994_g_at	MYL6B;MYL6;SMARCC
L2695:f13413;L3044:f10113;L3289:f10113;Welsh:39366_at;Uma:39366_at;Singh:39366_at	PPP1R3C
L2695:f15040;L3044:f14488;L3289:f14488;Uma:38408_at;Singh:38408_at	TSPAN7
L2695:f4811;L3044:f3;L3289:f3;Welsh:1035_g_at;Uma:1034_at	SYN3
L2695:f364;L3044:f3884;L3289:f3884;Uma:1897_at;Singh:1897_at	TGFBR3
L2695:f32280;L3289:f39605;Welsh:36956_at;Uma:36956_at;Singh:1137_at	SLC20A2
L2695:f3215;L3044:f1883;L3289:f1883;Welsh:38057_at;Uma:38057_at;Singh:38059_g_at	DPT
L2695:f10260;L3044:f9076;L3289:f9076;Welsh:32243_g_at;Uma:32243_g_at; Singh:32243_g_at	CRYAB;HSPB2;HSPB2-C11orf52;C11orf52
L2695:f2786;L3044:f6414;L3289:f6414;Welsh:32076_at;Uma:32076_at;Singh:32076_at	RCAN2
L2695:f35118;L3044:f16348;L3289:f16348;Welsh:32700_at;Uma:32700_at;Singh:32700_at	GBP2
L2695:f19545;L3044:f18241;L3289:f18241;Welsh:41744_at;Uma:41743_i_at;Singh:41744_at	OPTN
L2695:f25581;L3044:f13632;L3289:f24386;Welsh:40069_at;Uma:40069_at;Singh:40069_at	SVIL

The row names for the heatmap given in Figure 4.3.(continued)

Combination of probes	Symbol
L3044:f14657;L3289:f14657;Welsh:39691_at;Uma:39691_at;Singh:39691_at	SH3GLB1
L2695:f8705;L3044:f12277;L3289:f12277;Welsh:837_s_at;Uma:837_s_at;Singh:837_s_at	ME1
L2695:f3820;L3044:f4712;L3289:f4712;Welsh:32526_at;Uma:41862_at;Singh:32526_at	JAM3;NCAPD3; VPS26B;THYN1; ACAD8
L2695:f18875;L3044:f14579;Welsh:33924_at;Uma:33924_at;Singh:33924_at	DENND5A
L2695:f11944;L3044:f11696;L3289:f11041;Welsh:2056_at;Singh:2056_at	FGFR1
L2695:f5192;L3044:f3760;L3289:f3760;Welsh:31845_at;Singh:31845_at	ELF4
L2695:f3249;L3044:f423;L3289:f423;Welsh:32851_at;Uma:34532_at;Singh:32851_at	CELF2
L2695:f9562;L3044:f31036;L3289:f31036;Welsh:41870_at;Uma:31617_at;Singh:41871_at	PDPN
L2695:f39352;L3044:f38088;L3289:f38088;Uma:41802_at;Singh:40938_at	C11orf80
L2695:f23246;L3044:f38922;L3289:f30147;Welsh:35776_at;Uma:39096_at;Singh:488_at	GART;SON; DON-SON;CRYZL1; ITSN1
L2695:f37357;L3044:f40761;L3289:f40761;Welsh:39388_at;Uma:32104_i_at;Singh:650_s_at	KIAA0913
L2695:f29095;L3044:f36679;L3289:f24787;Welsh:39582_at;Singh:39582_at	CYLD
L2695:f5080;L3044:f30807;L3289:f30807;Welsh:33716_at;Singh:33716_at	PPP4R1L
L3044:f34438;L3289:f34438;Welsh:41719_i_at;Uma:39373_at;Singh:41720_r_at	FADS1
L2695:f3206;L3044:f2934;Welsh:38908_s_at;Uma:38908_s_at;Singh:36139_at	REV3L
L2695:f33091;L3044:f35135;L3289:f35135;Welsh:35737_at;Uma:35738_at;Singh:35737_at	HMGN4
L2695:f41168;L3044:f40780;L3289:f40780;Welsh:31791_at;Singh:31791_at	TP63
L2695:f2925;L3044:f2317;L3289:f2317;Welsh:41618_at;Singh:41618_at	COL17A1
L2695:f10399;L3044:f9987;L3289:f9987;Welsh:1898_at;Singh:1898_at	TRIM29
L2695:f16577;L3044:f16189;L3289:f16189;Welsh:40511_at;Singh:40511_at	GATA3
L2695:f6372;L3044:f38719;L3289:f38719;Welsh:1909_at;Singh:1909_at	BCL2
L2695:f41638;L3044:f40494;L3289:f40494;Welsh:862_at;Singh:862_at	SERPINB5
L2695:f3640;L3044:f3580;L3289:f3580;Welsh:34427_g_at;Singh:34425_at	MR1
L2695:f10048;L3289:f39833;Welsh:38312_at;Uma:38312_at;Singh:38312_at	OLFML2A
L2695:f24202;L3044:f23826;L3289:f23826;Welsh:40240_at;Singh:40240_at	GPRC5B
L2695:f43649;L3044:f36320;L3289:f36320;Welsh:41536_at;Singh:41536_at	ID4
L2695:f34645;L3044:f33929;L3289:f33929;Welsh:39315_at;Singh:39315_at	ANGPT1
L2695:f7106;L3044:f13664;L3289:f13664;Welsh:39054_at;Singh:39054_at	GSTM1
L2695:f39020;L3044:f42424;L3289:f42424;Welsh:556_s_at;Singh:556_s_at	GSTM4
L2695:f32138;L3044:f30824;L3289:f34762;Welsh:38044_at;Singh:38044_at	FAM107A
L2695:f38851;L3044:f37791;L3289:f19016;Welsh:1262_s_at;Singh:1262_s_at	TGFB2
L2695:f7621;L3044:f7573;L3289:f7573;Welsh:35277_at;Singh:35277_at	SPON1
L2695:f39024;L3044:f42428;L3289:f42428;Welsh:34818_at;Singh:34818_at	ETV5
L2695:f38194;L3044:f39970;L3289:f39970;Welsh:36521_at;Singh:36521_at	CLDN10
L2695:f13011;L3044:f9967;L3289:f9967;Welsh:1319_at;Singh:1319_at	DDR2
L2695:f37165;L3044:f40625;L3289:f40625;Welsh:1593_at;Singh:41806_at	FGF2
L2695:f39023;L3044:f42427;L3289:f24436;Welsh:40193_at;Singh:40193_at	LRRC23
L2695:f40335;L3044:f14353;L3289:f14353;Welsh:32224_at;Singh:32224_at	FCHSD2
L2695:f12272;L3044:f7364;L3289:f7364;Welsh:32694_at;Singh:32694_at	RARB
L2695:f4655;L3289:f37548;Welsh:37225_at;Uma:37225_at;Singh:37225_at	KANK1
L2695:f16895;L3044:f16287;L3289:f16287;Welsh:36119_at;Singh:36119_at	CAV1
L2695:f1488;L3289:f4532;Welsh:32102_at;Uma:32102_at;Singh:32102_at	SACS
L2695:f35492;L3044:f5798;L3289:f2929;Welsh:39673_i_at;Singh:39673_i_at	NOL8
L2695:f12488;L3044:f12028;L3289:f12028;Welsh:39750_at;Singh:39750_at	PARVA
L2695:f6689;L3044:f2925;L3289:f2925;Welsh:37716_at;Singh:37716_at	CD200
L2695:f2267;L3044:f32879;L3289:f32879;Welsh:39852_at;Uma:39852_at;Singh:39852_at	SPG20
L2695:f39171;L3044:f38171;L3289:f38171;Welsh:1336_s_at;Singh:160029_at	PRKCB
L2695:f16003;L3044:f19215;L3289:f19215;Welsh:41289_at;Singh:41289_at	NCAM1
L2695:f27244;L3044:f21940;L3289:f21940;Welsh:41385_at;Uma:41385_at;Singh:41385_at	EPB41L3
L2695:f5752;L3044:f5416;L3289:f5416;Welsh:34377_at;Singh:34377_at	ATP1A2

The row names for the heatmap given in Figure 4.3.(continued)

Combination of probes	Symbol
L2695:f18723;L3044:f15567;L3289:f15567;Welsh:1290_g_at;Singh:1290_g_at	GSTM5
L2695:f40851;L3044:f36967;L3289:f36967;Welsh:120_at;Singh:120_at	PELO
L2695:f18714;L3044:f39692;L3289:f39692;Welsh:33222_at;Singh:33222_at	FZD7
L3044:f34216;L3289:f34216;Welsh:38298_at;Uma:38298_at;Singh:38298_at	KCNIP1
L2695:f9313;L3044:f12601;L3289:f12601;Welsh:38098_at;Uma:38098_at;Singh:38098_at	LPIN1
L2695:f12413;L3044:f21595;L3289:f8249;Welsh:38722_at;Uma:38722_at;Singh:38722_at	COL6A1
L2695:f38361;L3044:f36964;L3289:f36964;Welsh:1276_g_at;Uma:1276_g_at;Singh:38047_at	RBPMS
L2695:f12638;L3044:f18141;L3289:f18141;Welsh:33862_at;Uma:33862_at;Singh:33862_at	PPAP2B
L2695:f19611;L3044:f26937;L3289:f26937;Welsh:32582_at;Uma:32582_at;Singh:774_g_at	KIAA0430
L2695:f24131;L3044:f23547;L3289:f23547;Welsh:37326_at;Uma:37326_at;Singh:37326_at	PLP2
L2695:f17538;L3044:f34268;L3289:f34268;Welsh:1734_at;Uma:1767_s_at;Singh:1767_s_at	TGFB3
L2695:f41051;L3044:f40727;L3289:f40727;Welsh:41449_at;Uma:41449_at	SGCE
L2695:f43817;L3044:f39061;L3289:f39061;Welsh:38415_at;Uma:38415_at	PTP4A2
L2695:f17536;L3044:f17088;L3289:f17088;Welsh:39939_at;Uma:39939_at;Singh:39939_at	PSMD10
L2695:f20532;L3044:f19996;L3289:f19996;Welsh:41143_at;Uma:41288_at;Singh:41144_g_at	CALM1
L2695:f25206;L3044:f27374;L3289:f27374;Welsh:40832_s_at;Uma:40832_s_at	TOR1AIP1
L2695:f29003;L3044:f16449;L3289:f16449;Welsh:35263_at;Singh:35263_at	EIF4EBP2
L2695:f16724;L3044:f20864;Welsh:32798_at;Uma:32798_at;Singh:32798_at	GSTM3
L2695:f37131;L3044:f39659;L3289:f39659;Welsh:39243_s_at;Uma:39243_s_at; Singh:37622_r_at	PSIP1
L2695:f13471;L3044:f35970;L3289:f35970;Welsh:35213_at;Uma:35213_at;Singh:35213_at	WBP4
L2695:f18241;L3044:f18118;L3289:f18118;Welsh:34843_at;Singh:34843_at	ZNF516
L2695:f6522;L3289:f12610;Welsh:38063_at;Uma:38118_at;Singh:37347_at	PBXIP1
L2695:f22551;L3044:f43119;L3289:f43119;Welsh:39878_at;Singh:39878_at	PCDH9
L2695:f38073;L3044:f12329;L3289:f37021;Welsh:38220_at;Singh:38220_at	DPYD
L2695:f30252;L3044:f23562;L3289:f23562;Welsh:34992_g_at;Singh:34993_at	SGCD
L2695:f17610;L3044:f39736;L3289:f39736;Welsh:41346_at;Singh:41346_at	LARGE
L2695:f25104;L3044:f33325;L3289:f27980;Welsh:39701_at;Uma:39701_at;Singh:39701_at	ZIM2
L2695:f28978;L3044:f28334;L3289:f28334;Welsh:39550_at;Uma:39550_at;Singh:39550_at	GLT25D2
L2695:f121;L3044:f27092;L3289:f27092;Welsh:40099_at;Uma:40099_at;Singh:40099_at	ARHGEF2
L3044:f21667;L3289:f21667;Welsh:31855_at;Uma:31855_at;Singh:31855_at	SRPX
L3044:f20343;L3289:f20343;Welsh:37589_at;Uma:37589_at;Singh:37590_g_at	ZNF710
L2695:f22893;L3289:f2085;Welsh:37924_g_at;Uma:37925_r_at;Singh:37924_g_at	APOM
L2695:f3360;L3044:f24302;Welsh:38175_at;Uma:38175_at;Singh:38175_at	CAND2
L2695:f5907;L3044:f6655;L3289:f41993;Uma:41549_s_at;Singh:41549_s_at	AP1S2
L2695:f20631;L3044:f115;L3289:f20923;Welsh:38045_at;Uma:38045_at	CTNND2
L2695:f499;L3044:f30866;L3289:f30866;Welsh:39602_at;Uma:39602_at	MYRIP
L2695:f38469;L3044:f36541;L3289:f26587;Welsh:34347_at;Singh:36624_at	P4HTM
L2695:f34299;L3289:f6218;Welsh:1805_g_at;Uma:40794_at;Singh:40794_at	KLK3
L2695:f4266;L3044:f130;L3289:f130;Welsh:40435_at;Uma:40435_at;Singh:40435_at	SLC25A6
L2695:f20346;L3289:f40664;Welsh:34608_at;Uma:34608_at;Singh:34608_at	GNB2L1
L2695:f37436;L3044:f39828;L3289:f39828;Welsh:31568_at;Uma:31568_at;Singh:31568_at	RPS10-NUDT3
L2695:f3894;L3044:f6337;L3289:f3614;Welsh:31545_at;Uma:31545_at;Singh:31545_at	VPS52
L2695:f24;L3044:f4668;L3289:f4668;Welsh:36587_at;Uma:36587_at;Singh:36587_at	EEF2
L2695:f42665;L3044:f37717;L3289:f37717;Uma:31583_at;Singh:31583_at	RPS8
L2695:f37167;L3044:f40627;L3289:f40627;Welsh:374_f_at;Uma:41181_r_at; Singh:33689_s_at	DDTL
L2695:f22226;L3044:f10980;L3289:f10980;Uma:38482_at;Singh:38482_at	CTDNEP1
L2695:f14963;L3044:f37009;L3289:f37009;Welsh:39817_s_at;Singh:39817_s_at	C6orf108
L2695:f23219;L3289:f21659;Welsh:38429_at;Uma:38429_at;Singh:38429_at	STRA13
L2695:f21536;L3044:f20788;L3289:f20788;Welsh:38464_at;Uma:38464_at;Singh:38464_at	INO80B
L2695:f3895;L3044:f3615;L3289:f3615;Welsh:1637_at;Singh:1637_at	MAPKAPK3
L2695:f44074;L3044:f16527;L3289:f41754;Welsh:39748_at;Uma:39748_at;Singh:39748_at	SLC7A1
L2695:f35455;L3044:f34287;L3289:f34287;Welsh:39598_at;Uma:39598_at;Singh:38614_s_at	BCYRN1

The row names for the heatmap given in Figure 4.3.(continued)

Combination of probes	Symbol
L2695:f6678;L3044:f6314;L3289:f6314;Welsh:37193_at;Singh:37193_at	UCK2
L2695:f16333;L3044:f18929;L3289:f18929;Welsh:32001_s_at;Uma:32001_s_at; Singh:32001_s_at	PCSK6
L2695:f39246;L3289:f39122;Welsh:37068_at;Uma:37068_at;Singh:37068_at	TDRD6
L2695:f6539;L3289:f6439;Welsh:36936_at;Uma:36936_at;Singh:36936_at	TSTA3
L2695:f18747;L3044:f28042;L3289:f5337;Uma:41642_at;Singh:40553_at	LOC100170939
L2695:f18747;L3044:f5337;L3289:f5337;Uma:40552_s_at;Singh:40553_at	GUSBP3
L2695:f6687;L3044:f5337;L3289:f5337;Uma:41642_at;Singh:40553_at	SMA5
L2695:f6687;L3044:f5337;L3289:f5337;Uma:40552_s_at;Singh:40553_at	GUSBP9
L2695:f26;L3044:f32643;L3289:f32643;Welsh:40176_at;Singh:40176_at	TRIM27
L2695:f25379;L3044:f38878;L3289:f38878;Welsh:33126_at;Uma:33126_at	PBRM1
L2695:f33094;L3044:f5482;L3289:f35138;Welsh:36484_at;Uma:36483_at;Singh:36483_at	GALNT3
L2695:f35369;L3044:f35309;L3289:f35309;Welsh:32634_s_at;Uma:32634_s_at; Singh:32634_s_at	ICA1
L2695:f19050;L3044:f31635;L3289:f35237;Welsh:31600_s_at;Singh:41250_at	PMS2
L2695:f34355;L3044:f28915;L3289:f28915;Welsh:39396_at;Singh:39396_at	LYPLA1
L2695:f34360;L3044:f11050;L3289:f28920;Welsh:36654_s_at;Singh:38085_at	HNRNPA2B1
L2695:f7354;L3289:f3782;Welsh:32051_at;Uma:32051_at;Singh:32051_at	ALG8
L2695:f7816;L3289:f8324;Welsh:35835_at;Uma:35835_at;Singh:35835_at	ANAPC5
L2695:f30148;L3044:f29068;L3289:f29068;Welsh:38716_at;Uma:38716_at;Singh:38716_at	CAMKK2
L2695:f26255;L3044:f17467;L3289:f40139;Welsh:32684_at;Uma:32684_at;Singh:32684_at	C9orf91
L2695:f28778;L3044:f39224;L3289:f39224;Welsh:36160_s_at;Uma:36160_s_at; Singh:36160_s_at	PTPRN2
L2695:f40064;L3044:f16102;L3289:f16102;Welsh:41468_at;Uma:41468_at;Singh:41468_at	TARP
L2695:f3928;L3044:f7792;L3289:f7792;Welsh:37366_at;Uma:37366_at;Singh:40059_r_at	PDLIM5
L2695:f40766;L3044:f26446;L3289:f26446;Welsh:33358_at;Uma:33358_at;Singh:33358_at	PPM1H
L2695:f11412;L3044:f12464;L3289:f12464;Welsh:34050_at;Uma:34050_at;Singh:34050_at	ACSM1
L2695:f26968;L3044:f25344;L3289:f25344;Welsh:37639_at;Uma:37639_at;Singh:37639_at	SCN1B
L2695:f37438;L3044:f39830;L3289:f39830;Welsh:41706_at;Uma:41706_at;Singh:41706_at	C1QTNF3-AMACR
L2695:f4277;L3044:f22116;L3289:f22116;Welsh:40548_at;Uma:40548_at	BICD1
L2695:f2701;L3044:f31476;L3289:f31476;Welsh:39925_at;Uma:39925_at;Singh:39925_at	COL9A2
L2695:f44306;L3044:f7122;L3289:f37581;Welsh:38218_at;Uma:38218_at;Singh:38218_at	GCNT1
L2695:f7986;L3044:f7122;L3289:f3526;Welsh:35071_s_at;Singh:35071_s_at	GMDS
L2695:f41831;L3289:f37139;Welsh:39394_at;Uma:39394_at;Singh:39394_at	C2orf72
L2695:f26422;L3289:f28793;Welsh:1585_at;Uma:1585_at;Singh:1585_at	ERBB3
L2695:f39081;L3044:f38041;Welsh:37117_at;Uma:37117_at;Singh:37117_at	PRR5
L2695:f7664;L3044:f11964;Welsh:32444_at;Uma:41077_at;Singh:32444_at	ZNF862
L2695:f43833;L3289:f15917;Welsh:40982_at;Uma:36129_at;Singh:36129_at	HIC1
L2695:f6960;L3044:f17475;L3289:f17475;Welsh:32723_at;Uma:34851_at	C20orf108
L2695:f15006;L3044:f18342;L3289:f13629;Welsh:33131_at;Uma:33131_at;Singh:33131_at	SOX4
L2695:f2993;L3044:f11027;L3289:f18766;Welsh:32124_at;Uma:32124_at	THOC2
L2695:f20946;L3044:f35553;L3289:f35553;Welsh:35214_at;Uma:35214_at;Singh:35214_at	UGDH
L2695:f1260;L3044:f1256;L3289:f1256;Uma:914_g_at;Singh:2046_at	ERG
L2695:f32664;L3044:f27309;L3289:f27309;Welsh:33375_at;Uma:33375_at;Singh:33375_at	MYO6
L2695:f37388;L3044:f37320;L3289:f37320;Uma:38803_at;Singh:38803_at	NCALD
L2695:f31570;L3044:f8994;L3289:f35262;Welsh:39732_at;Uma:39732_at;Singh:39732_at	MAP7
L2695:f25078;L3044:f23462;L3289:f23462;Welsh:34327_at;Uma:34327_at	HLTF
L2695:f21768;L3044:f29386;L3289:f29386;Welsh:41242_at;Singh:41242_at	UAP1
L2695:f188;L3044:f84;L3289:f84;Welsh:1969_s_at;Singh:33317_at	CDK7
L2695:f41979;L3044:f36211;L3289:f36211;Welsh:39542_at;Uma:39542_at;Singh:39542_at	ENC1
L2695:f7291;L3044:f14964;L3289:f2691;Welsh:32034_at;Uma:32034_at;Singh:32034_at	ZNF217
L2695:f11222;L3044:f13054;L3289:f8642;Welsh:36918_at;Uma:36918_at;Singh:36918_at	GUCY1A3
L2695:f24046;L3289:f35963;Welsh:39608_at;Uma:39609_at;Singh:39608_at	SIM2
L2695:f12646;L3289:f8474;Welsh:33267_at;Uma:33267_at;Singh:33267_at	ATP8A1

The row names for the heatmap given in Figure 4.3.(continued)

Combination of probes	Symbol
L2695:f14927;L3044:f14831;L3289:f14831;Welsh:36965_at;Uma:36967_g_at;Singh:36965_at	ANK3
L2695:f27253;L3289:f2874;Welsh:40576_f_at;Uma:32393_s_at;Singh:32393_s_at	HNRPDL
L2695:f33780;L3289:f15900;Welsh:35829_at;Uma:35829_at;Singh:35829_at	CADM1
L3044:f31685;L3289:f3164;Welsh:32203_at;Uma:32203_at;Singh:32203_at	RBCK1
L2695:f9491;L3044:f9251;L3289:f9251;Welsh:37955_at;Uma:37955_at;Singh:37955_at	CNPY2
L2695:f878;L3289:f806;Welsh:32113_at;Uma:32113_at;Singh:32112_s_at	AIM1
L2695:f13032;L3044:f6712;L3289:f29947;Welsh:40765_at;Singh:40765_at	PDXDC1
L2695:f34629;L3289:f29365;Welsh:34840_at;Uma:34840_at;Singh:34840_at	SERINC5

Combination of probes is the list of combined probes. **Symbol** is the list of gene symbols corresponding to each combined probes.

The row names for the heatmap given in Figure 4.4

Table 8.4: **The row names for the heatmap given in Figure 4.4.**

Combination of probes	Symbol
L2695:f35369;L3044:f35309;L3289:f35309;Welsh:32634_s_at;Uma:32634_s_at; Singh:32634_s_at	ICA1
L2695:f33094;L3044:f5482;L3289:f35138;Welsh:36484_at;Uma:36483_at;Singh:36483_at	GALNT3
L2695:f16333;L3044:f18929;L3289:f18929;Welsh:32001_s_at;Uma:32001_s_at; Singh:32001_s_at	PCSK6
L2695:f35455;L3044:f34287;L3289:f34287;Welsh:39598_at;Uma:39598_at; Singh:38614_s_at	BCYRN1;GJB1; ZMYM3;NONO; ITGB1BP2;TAF1; INGX;OGT;ACRC; CXCR3; LOC100132741; FLJ46446;CXorf49B; CXorf49
L2695:f44074;L3044:f16527;L3289:f41754;Welsh:39748_at;Uma:39748_at;Singh:39748_at	SLC7A1
L2695:f31570;L3044:f8994;L3289:f35262;Welsh:39732_at;Uma:39732_at;Singh:39732_at	MAP7
L2695:f32664;L3044:f27309;L3289:f27309;Welsh:33375_at;Uma:33375_at;Singh:33375_at	MYO6
L2695:f2701;L3044:f31476;L3289:f31476;Welsh:39925_at;Uma:39925_at;Singh:39925_at	COL9A2
L2695:f44306;L3044:f7122;L3289:f37581;Welsh:38218_at;Uma:38218_at;Singh:38218_at	GCNT1
L2695:f37438;L3044:f39830;L3289:f39830;Welsh:41706_at;Uma:41706_at;Singh:41706_at	C1QTNF3-AMACR; AMACR;C1QTNF3
L2695:f11412;L3044:f12464;L3289:f12464;Welsh:34050_at;Uma:34050_at;Singh:34050_at	ACSM1
L2695:f14927;L3044:f14831;L3289:f14831;Welsh:36965_at;Uma:36967_g_at; Singh:36965_at	ANK3
L2695:f20946;L3044:f35553;L3289:f35553;Welsh:35214_at;Uma:35214_at;Singh:35214_at	UGDH
L2695:f15006;L3044:f18342;L3289:f13629;Welsh:33131_at;Uma:33131_at;Singh:33131_at	SOX4
L2695:f7291;L3044:f14964;L3289:f2691;Welsh:32034_at;Uma:32034_at;Singh:32034_at	ZNF217
L2695:f41979;L3044:f36211;L3289:f36211;Welsh:39542_at;Uma:39542_at;Singh:39542_at	ENC1
L2695:f11222;L3044:f13054;L3289:f8642;Welsh:36918_at;Uma:36918_at;Singh:36918_at	GUCY1A3
L2695:f40766;L3044:f26446;L3289:f26446;Welsh:33358_at;Uma:33358_at;Singh:33358_at	PPM1H
L2695:f3928;L3044:f7792;L3289:f7792;Welsh:37366_at;Uma:37366_at;Singh:40059_r_at	PDLIM5
L2695:f40064;L3044:f16102;L3289:f16102;Welsh:41468_at;Uma:41468_at;Singh:41468_at	TARP
L2695:f28778;L3044:f39224;L3289:f39224;Welsh:36160_s_at;Uma:36160_s_at; Singh:36160_s_at	PTPRN2;MIR153-2; LOC100506585; MIR595

The row names for the heatmap given in Figure 4.4.(continued)

Combination of probes	Symbol
L2695:f26255;L3044:f17467;L3289:f40139;Welsh:32684_at;Uma:32684_at;Singh:32684_at	C9orf91
L2695:f30148;L3044:f29068;L3289:f29068;Welsh:38716_at;Uma:38716_at;Singh:38716_at	CAMKK2
L2695:f26968;L3044:f25344;L3289:f25344;Welsh:37639_at;Uma:37639_at;Singh:37639_at	SCN1B;HPN; LOC100128675
L2695:f21536;L3044:f20788;L3289:f20788;Welsh:38464_at;Uma:38464_at;Singh:38464_at	INO80B; INO80B-WBP1; WBP1;MOGS
L2695:f9491;L3044:f9251;L3289:f9251;Welsh:37955_at;Uma:37955_at;Singh:37955_at	CNPY2;PAN2
L2695:f37167;L3044:f40627;L3289:f40627;Welsh:374_f_at;Uma:41181_r_at; Singh:33689_s_at	DDTL;DDT;GSTT2
L2695:f24;L3044:f4668;L3289:f4668;Welsh:36587_at;Uma:36587_at;Singh:36587_at	EEF2;SNORD37
L2695:f4266;L3044:f130;L3289:f130;Welsh:40435_at;Uma:40435_at;Singh:40435_at	SLC25A6
L2695:f3894;L3044:f6337;L3289:f3614;Welsh:31545_at;Uma:31545_at;Singh:31545_at	VPS52;RPS18; B3GALT4;WDR46; PFDN6;RGL2; TAPBP;ZBTB22; DAXX
L2695:f37436;L3044:f39828;L3289:f39828;Welsh:31568_at;Uma:31568_at;Singh:31568_at	RPS10-NUDT3;NUDT3; RPS10
L2695:f16672;L3044:f19748;L3289:f19748;Welsh:33820_g_at;Uma:33819_at; Singh:33820_g_at	LDHB
L2695:f25093;L3044:f11703;L3289:f11703;Welsh:33994_g_at;Uma:38251_at; Singh:33994_g_at	MYL6B;MYL6; SMARCC2
L2695:f6668;L3044:f1804;L3289:f1804;Welsh:32314_g_at;Uma:32314_g_at; Singh:32314_g_at	CA9
L2695:f10260;L3044:f9076;L3289:f9076;Welsh:32243_g_at;Uma:32243_g_at; Singh:32243_g_at	CRYAB;HSPB2; HSPB2-C11orf52; C11orf52
L2695:f3215;L3044:f1883;L3289:f1883;Welsh:38057_at;Uma:38057_at;Singh:38059_g_at	DPT
L2695:f17538;L3044:f34268;L3289:f34268;Welsh:1734_at;Uma:1767_s_at; Singh:1767_s_at	TGFB3
L2695:f24131;L3044:f23547;L3289:f23547;Welsh:37326_at;Uma:37326_at;Singh:37326_at	PLP2;PRICKLE3
L2695:f37082;L3044:f32334;L3289:f32334;Welsh:769_s_at;Uma:757_at;Singh:769_s_at	ANXA2
L2695:f35118;L3044:f16348;L3289:f16348;Welsh:32700_at;Uma:32700_at;Singh:32700_at	GBP2
L2695:f27244;L3044:f21940;L3289:f21940;Welsh:41385_at;Uma:41385_at;Singh:41385_at	EPB41L3
L2695:f2267;L3044:f32879;L3289:f32879;Welsh:39852_at;Uma:39852_at;Singh:39852_at	SPG20
L2695:f121;L3044:f27092;L3289:f27092;Welsh:40099_at;Uma:40099_at;Singh:40099_at	ARHGEF2
L2695:f25104;L3044:f33325;L3289:f27980;Welsh:39701_at;Uma:39701_at;Singh:39701_at	ZIM2;PEG3; PEG3-AS1;MIMT1
L2695:f28978;L3044:f28334;L3289:f28334;Welsh:39550_at;Uma:39550_at;Singh:39550_at	GLT25D2
L2695:f13471;L3044:f35970;L3289:f35970;Welsh:35213_at;Uma:35213_at;Singh:35213_at	WBP4
L2695:f8705;L3044:f12277;L3289:f12277;Welsh:837_s_at;Uma:837_s_at;Singh:837_s_at	ME1
L2695:f37131;L3044:f39659;L3289:f39659;Welsh:39243_s_at;Uma:39243_s_at; Singh:37622_r_at	PSIP1
L2695:f38361;L3044:f36964;L3289:f36964;Welsh:1276_g_at;Uma:1276_g_at; Singh:38047_at	RBPMS
L2695:f9313;L3044:f12601;L3289:f12601;Welsh:38098_at;Uma:38098_at;Singh:38098_at	LPIN1
L2695:f12413;L3044:f21595;L3289:f8249;Welsh:38722_at;Uma:38722_at;Singh:38722_at	COL6A1
L2695:f12638;L3044:f18141;L3289:f18141;Welsh:33862_at;Uma:33862_at;Singh:33862_at	PPAP2B
L2695:f19611;L3044:f26937;L3289:f26937;Welsh:32582_at;Uma:32582_at;Singh:774_g_at	KIAA0430;NDE1; MIR484;MYH11
L2695:f25581;L3044:f13632;L3289:f24386;Welsh:40069_at;Uma:40069_at;Singh:40069_at	SVIL;MIR604; MIR938

The row names for the heatmap given in Figure 4.4.(continued)

Combination of probes	Symbol
L2695:f19545;L3044:f18241;L3289:f18241;Welsh:41744_at;Uma:41743_i_at; Singh:41744_at	OPTN
L2695:f3820;L3044:f4712;L3289:f4712;Welsh:32526_at;Uma:41862_at;Singh:32526_at	JAM3;NCAPD3; VPS26B;THYN1; ACAD8
L2695:f3249;L3044:f423;L3289:f423;Welsh:32851_at;Uma:34532_at;Singh:32851_at	CELF2
L2695:f9562;L3044:f31036;L3289:f31036;Welsh:41870_at;Uma:31617_at;Singh:41871_at	PDPN
L2695:f23246;L3044:f38922;L3289:f30147;Welsh:35776_at;Uma:39096_at;Singh:488_at	GART;SON; DON-SON;CRYZL1; ITSN1
L2695:f37357;L3044:f40761;L3289:f40761;Welsh:39388_at;Uma:32104_i_at; Singh:650_s_at	KIAA0913; LOC100507331; NDST2;CAMK2G
L2695:f33091;L3044:f35135;L3289:f35135;Welsh:35737_at;Uma:35738_at;Singh:35737_at	HMGN4
L2695:f20532;L3044:f19996;L3289:f19996;Welsh:41143_at;Uma:41288_at; Singh:41144_g_at	CALM1
L2695:f17536;L3044:f17088;L3289:f17088;Welsh:39939_at;Uma:39939_at;Singh:39939_at	PSMD10;ATG4A; COL4A6;COL4A5
L2695:f13413;L3044:f10113;L3289:f10113;Welsh:39366_at;Uma:39366_at;Singh:39366_at	PPP1R3C
L2695:f2786;L3044:f6414;L3289:f6414;Welsh:32076_at;Uma:32076_at;Singh:32076_at	RCAN2

Combination of probes is the list of combined probes. **Symbol** is the list of gene symbols corresponding to each combined probes.

Comparison result of RankProd and Coloured (α, β)-k Feature Set problem approach.

Table 8.5: **The camparison result of Coloured (α, β)-k Feature Set problem approach and RankProd.**

Symbol	RP/RSum	p-value	CABk-DS	RP
HPN	227.9175	2.00E-04	Six	Up
PDLIM5	228.9312	0.0053	Six	Up
PAN2	518.8578	0.0061	Six	Up
TARP	565.9352	0.0001	Six	Up
DDT	604.6938	0.0001	Six	Up
CELF2	656.1007	0.0001	Six	Up
SLC25A6	657.3755	0.0001	Six	Up
EEF2	668.7355	0.0001	Six	Up
DPT	695.8718	0.0004	Six	Up
GART	709.2517	0.0015	Six	Up
HSPB2	750.9518	0.004	Six	Up
RPS10	824.37	1.00E-04	Six	Up
PPM1H	842.2355	1.00E-04	Six	Up
GCNT1	941.0459	1.00E-04	Six	Up
CAMKK2	975.4789	1.00E-04	Six	Up
SOX4	1023.6392	6.00E-04	Six	Up
NONO	1028.5102	7.00E-04	Six	Up
COL9A2	1034.302	0.001	Six	Up
CNPY2	1052.7785	0.0015	Six	Up
GUCY1A3	1057.6461	0.0031	Six	Up

The camparison result of Coloured (α, β)-k Feature Set problem approach and RankProd. (continued)

Symbol	RP/RSum	p-value	CABk-DS	RP
UGDH	1057.9905	0.0035	Six	Up
ZNF217	1075.9191	0.0043	Six	Up
PTPRN2	1080.5918	0.0044	Six	Up
NCAPD3	1195.0817	0.0059	Six	Up
GJB1	1200.7698	0.0001	Six	Up
MYO6	1250.4656	0.0001	Six	Up
RGL2	1428.914	0.0001	Six	Up
ICA1	1492.0196	0.0001	Six	Up
GALNT3	1499.6988	0.0002	Six	Up
TAPBP	1508.5439	0.0003	Six	Up
C9orf91	1553.8305	0.0005	Six	Up
ENC1	1575.6187	0.0007	Six	Up
MOGS	1586.9525	0.0011	Six	Up
MAP7	1661.5172	0.002	Six	Up
PCSK6	1684.7031	0.0024	Six	Up
SLC7A1	1691.6255	0.0027	Six	Up
OGT	1769.0347	0.0043	Six	Up
ANK3	1976.3703	0.0057	Six	Up
SON	1987.4556	0.0059	Six	Up
CRYAB	452.5998	1.00E-04	Six	Down
SMARCC2	465.1049	1.00E-04	Six	Down
PFDN6	512.0754	1.00E-04	Six	Down
MYH11	564.612	1.00E-04	Six	Down
TPM2	574.0013	2.00E-04	Six	Down
VPS52	741.8925	4.00E-04	Six	Down
COL4A6	779.7312	8.00E-04	Six	Down
INGX	823.1692	0.0011	Six	Down
INO80B	850.91	0.003	Six	Down
LDHB	890.8802	0.0001	Six	Down
SVIL	891.2703	0.0001	Six	Down
GSTT2	967.5964	0.0001	Six	Down
KIAA0913	974.3261	0.0001	Six	Down
HMGN4	990.5274	0.0001	Six	Down
ITGB1BP2	998.1156	0.0001	Six	Down
ME1	1044.0996	0.0002	Six	Down
TGFB3	1137.3184	0.0002	Six	Down
RCAN2	1166.9677	0.0004	Six	Down
COL6A1	1197.2713	0.0005	Six	Down
PSIP1	1222.9958	0.0005	Six	Down
PPAP2B	1239.8578	0.0005	Six	Down
GLT25D2	1279.3628	0.0007	Six	Down
CALM1	1284.5084	0.0007	Six	Down
GBP2	1365.2355	0.001	Six	Down
RBPMS	1369.1305	0.0011	Six	Down
PLP2	1375.5231	0.0012	Six	Down
LPIN1	1429.9997	0.0045	Six	Down
MYL6	1459.2016	0.0054	Six	Down
SPG20	1461.7208	0.0062	Six	Down
JAM3	1491.8871	0.0068	Six	Down
OPTN	1556.2914	1.00E-04	Six	Down
PEG3	1624.7228	1.00E-04	Six	Down
CAMK2G	1715.7727	1.00E-04	Six	Down
ANXA2	1716.9316	1.00E-04	Six	Down
EPB41L3	1717.0527	1.00E-04	Six	Down

The camparison result of Coloured (α, β)-k Feature Set problem approach and RankProd. (continued)

Symbol	RP/RSum	p-value	CABk-DS	RP
ITSN1	1748.0398	1.00E-04	Six	Down
COL4A5	1755.6252	1.00E-04	Six	Down
MYL6B	1809.9137	1.00E-04	Six	Down
PDPN	1889.0452	1.00E-04	Six	Down
WBP4	1957.2202	1.00E-04	Six	Down
TSPAN1	373.5353	1.00E-04	Four	Up
FBP1	672.8917	1.00E-04	Four	Up
RPL6	683.5961	1.00E-04	Four	Up
RPS19	704.0116	1.00E-04	Four	Up
CLDN8	711.6421	1.00E-04	Four	Up
LRIG1	737.5682	1.00E-04	Four	Up
HOXB13	740.25	1.00E-04	Four	Up
RPL10	783.0239	1.00E-04	Four	Up
SERP1	793.1767	1.00E-04	Four	Up
RPL11	808.9098	1.00E-04	Four	Up
THBS4	858.0624	1.00E-04	Four	Up
RGS10	919.5606	1.00E-04	Four	Up
TMED3	965.0283	1.00E-04	Four	Up
APOC1	1002.7739	1.00E-04	Four	Up
CD2AP	1115.2168	1.00E-04	Four	Up
ARFIP2	1127.6229	1.00E-04	Four	Up
ATP5J2	1129.6675	1.00E-04	Four	Up
NDUFA2	1205.7477	1.00E-04	Four	Up
ATP2C1	1221.8625	1.00E-04	Four	Up
FMO5	1236.7299	1.00E-04	Four	Up
SFPQ	1243.7394	1.00E-04	Four	Up
BOLA2	1266.3046	1.00E-04	Four	Up
IQGAP2	1312.9738	1.00E-04	Four	Up
SMARCA4	1343.8239	1.00E-04	Four	Up
SLC43A1	1344.3198	1.00E-04	Four	Up
ST14	1409.9499	1.00E-04	Four	Up
POLR2H	1470.1269	1.00E-04	Four	Up
TAOK3	1510.5051	1.00E-04	Four	Up
APRT	1513.3514	1.00E-04	Four	Up
TMEM106C	1530.4547	1.00E-04	Four	Up
PDE3B	1537.4501	1.00E-04	Four	Up
CHN2	1612.6078	1.00E-04	Four	Up
ACTL6A	1671.793	1.00E-04	Four	Up
STIL	1675.1578	1.00E-04	Four	Up
HOMER2	1686.8916	2.00E-04	Four	Up
ATP7B	1721.8891	2.00E-04	Four	Up
MTHFD2	1738.4284	2.00E-04	Four	Up
CYFIP2	1740.9804	2.00E-04	Four	Up
FDPS	1747.9081	2.00E-04	Four	Up
PATZ1	1768.5346	2.00E-04	Four	Up
GNG5	1831.4393	2.00E-04	Four	Up
EIF2B4	1844.0263	2.00E-04	Four	Up
ACSM3	1863.4744	2.00E-04	Four	Up
POLB	1893.8188	2.00E-04	Four	Up
IMPDH1	1940.0713	2.00E-04	Four	Up
GOLGA2	1948.2577	2.00E-04	Four	Up
KIAA0100	1960.9115	2.00E-04	Four	Up
GPX1	1963.452	2.00E-04	Four	Up
BUD31	1984.5819	2.00E-04	Four	Up

The camparison result of Coloured (α, β)-k Feature Set problem approach and RankProd. (continued)

Symbol	RP/RSum	p-value	CABk-DS	RP
G0S2	874.5298	2.00E-04	Four	Down
WIF1	915.3743	2.00E-04	Four	Down
FXYD1	1054.1995	2.00E-04	Four	Down
COL4A1	1062.0875	2.00E-04	Four	Down
SERPINE2	1103.3635	2.00E-04	Four	Down
COL4A2	1108.5902	2.00E-04	Four	Down
FN1	1210.7953	2.00E-04	Four	Down
TGFB1I1	1228.4363	3.00E-04	Four	Down
PDGFRA	1262.0877	3.00E-04	Four	Down
TBL1X	1272.3462	3.00E-04	Four	Down
GJA1	1279.1386	3.00E-04	Four	Down
GPX2	1401.3992	3.00E-04	Four	Down
MGST3	1435.3482	3.00E-04	Four	Down
LAMA2	1436.254	3.00E-04	Four	Down
MME	1450.4809	3.00E-04	Four	Down
PARM1	1464.8974	3.00E-04	Four	Down
MTMR11	1500.733	3.00E-04	Four	Down
PLS3	1506.7552	3.00E-04	Four	Down
COL13A1	1521.0313	3.00E-04	Four	Down
RHOA	1569.114	3.00E-04	Four	Down
MEG3	1582.0559	3.00E-04	Four	Down
ACSL4	1590.8792	3.00E-04	Four	Down
NID2	1612.4671	3.00E-04	Four	Down
ACOX2	1624.3872	4.00E-04	Four	Down
STX2	1657.6756	4.00E-04	Four	Down
CDKN1C	1692.0311	4.00E-04	Four	Down
ECM2	1695.7078	4.00E-04	Four	Down
DOCK3	1703.8333	4.00E-04	Four	Down
SLC2A5	1708.7135	4.00E-04	Four	Down
GAS6	1723.0399	4.00E-04	Four	Down
IFI16	1729.693	4.00E-04	Four	Down
KCNJ8	1737.8417	4.00E-04	Four	Down
BCL11A	1768.4014	4.00E-04	Four	Down
SLC25A4	1786.072	4.00E-04	Four	Down
CHN1	1826.5428	4.00E-04	Four	Down
CDC42BPA	1842.4336	4.00E-04	Four	Down
FNTB	1872.5274	5.00E-04	Four	Down
PYGL	1910.2775	5.00E-04	Four	Down
BCAR3	1921.4493	5.00E-04	Four	Down
LSAMP	1929.446	5.00E-04	Four	Down
STAT5B	1962.3023	5.00E-04	Four	Down
FGFR4	1982.0775	5.00E-04	Four	Down
GBP1	1984.1519	5.00E-04	Four	Down
GNB2L1	493.1805	5.00E-04	Five	Up
KLK3	531.9704	5.00E-04	Five	Up
AIM1	625.4295	6.00E-04	Five	Up
PLA2G7	651.7188	6.00E-04	Five	Up
TMEM87A	789.0306	6.00E-04	Five	Up
RPS8	852.8789	6.00E-04	Five	Up
UAP1	859.5959	6.00E-04	Five	Up
IMPDH2	886.89	6.00E-04	Five	Up
NCALD	952.5339	6.00E-04	Five	Up
LYPLA1	979.8226	6.00E-04	Five	Up
CLDN7	1143.6496	6.00E-04	Five	Up

The camparison result of Coloured (α, β)-k Feature Set problem approach and RankProd. (continued)

Symbol	RP/RSum	p-value	CABk-DS	RP
BICD1	1152.6465	6.00E-04	Five	Up
STRA13	1158.2901	6.00E-04	Five	Up
ATP8A1	1176.9219	6.00E-04	Five	Up
CADM1	1279.4535	7.00E-04	Five	Up
ERG	1395.1241	7.00E-04	Five	Up
TSTA3	1440.1374	7.00E-04	Five	Up
CBX3	1469.5455	7.00E-04	Five	Up
PDXDC1	1576.2758	7.00E-04	Five	Up
GMDS	1619.8455	7.00E-04	Five	Up
C6orf108	1644.9673	8.00E-04	Five	Up
SIM2	1668.3561	8.00E-04	Five	Up
C2orf72	1726.7694	8.00E-04	Five	Up
CDK7	1748.1595	8.00E-04	Five	Up
ALG8	1761.0326	8.00E-04	Five	Up
ANAPC5	1765.9081	8.00E-04	Five	Up
AIMP2	1790.2041	9.00E-04	Five	Up
AURKA	1816.1947	9.00E-04	Five	Up
UCK2	1851.0831	9.00E-04	Five	Up
ERBB3	1873.134	9.00E-04	Five	Up
GLT8D1	1938.6486	9.00E-04	Five	Up
CAV1	550.3004	9.00E-04	Five	Down
TP63	937.3143	9.00E-04	Five	Down
PMP22	967.9335	9.00E-04	Five	Down
GPRC5B	1054.5881	9.00E-04	Five	Down
ANGPT1	1068.1195	9.00E-04	Five	Down
KCNMB1	1111.4402	9.00E-04	Five	Down
TGFBR3	1113.6512	9.00E-04	Five	Down
GSTM5	1121.004	0.001	Five	Down
SPON1	1147.2573	0.001	Five	Down
ATP1A2	1167.4275	0.001	Five	Down
CALD1	1192.4459	0.001	Five	Down
CD200	1254.6717	0.001	Five	Down
CAND2	1255.4921	0.001	Five	Down
NCAM1	1313.6473	0.001	Five	Down
GAS1	1335.8615	0.0011	Five	Down
KANK1	1343.8351	0.0011	Five	Down
FAM107A	1386.0465	0.0011	Five	Down
GSTM3	1393.4906	0.0011	Five	Down
TRIM29	1405.6703	0.0011	Five	Down
ENO2	1407.7495	0.0011	Five	Down
LRRC23	1419.5738	0.0011	Five	Down
GATA3	1432.0575	0.0011	Five	Down
FADS1	1448.0613	0.0012	Five	Down
SERPINB5	1464.0379	0.0012	Five	Down
FZD7	1484.8933	0.0012	Five	Down
PRKCB	1490.9054	0.0012	Five	Down
DDR2	1526.2323	0.0012	Five	Down
TSPAN7	1532.9795	0.0013	Five	Down
FCHSD2	1585.6345	0.0013	Five	Down
FGFR2	1593.4757	0.0013	Five	Down
COL17A1	1607.6265	0.0013	Five	Down
ETV5	1631.6247	0.0013	Five	Down
SACS	1635.9795	0.0013	Five	Down
PLAGL1	1670.1165	0.0013	Five	Down

The camparison result of Coloured (α, β)-k Feature Set problem approach and RankProd. (continued)

Symbol	RP/RSum	p-value	CABk-DS	RP
SH3GLB1	1682.1846	0.0013	Five	Down
OLFML2A	1682.9064	0.0014	Five	Down
TIMP3	1692.4517	0.0014	Five	Down
ZNF516	1708.3773	0.0014	Five	Down
TUBA4A	1737.6001	0.0014	Five	Down
AP1S2	1737.7166	0.0014	Five	Down
ID4	1740.3777	0.0014	Five	Down
TOR1AIP1	1747.5347	0.0014	Five	Down
SRPX	1761.1132	0.0014	Five	Down
FGFR1	1775.6016	0.0014	Five	Down
PTP4A2	1782.3078	0.0015	Five	Down
FGF2	1807.5369	0.0015	Five	Down
MR1	1821.5117	0.0015	Five	Down
RARB	1824.9445	0.0015	Five	Down
REV3L	1826.9822	0.0016	Five	Down
EIF4EBP2	1839.6907	0.0016	Five	Down
CYLD	1842.1145	0.0016	Five	Down
ST6GALNAC2	1866.3676	0.0017	Five	Down
BCL2	1872.4913	0.0017	Five	Down
PEG10	1875.2925	0.0018	Five	Down
DZIP1	1965.6731	0.0018	Five	Down
DENND5A	1977.7398	0.0019	Five	Down
RAP1A	1987.8412	0.0019	Five	Down
SHC1	1994.9624	0.0019	Five	Down

Symbol is the list of gene symbols that are resulted from RankProd. **RP/RSum** is the rank product resulted from RankProd. **p-value** is the resulted p-value from RankProd. **CABk-DS** is the number of dataset in which that particular gene is present according to the Coloured (α, β)-k Feature Set problem approach result.**RP** indicates that the gene is up or down regulated in RankProd result

Functional analysis result of Coloured (α, β)-k Feature Set problem approach resulted genes.

Table 8.6: Functional analysis result of Coloured (α, β)-k Feature Set problem approach resulted genes.

Group	Symbol	Score
G1	MYL6;SVIL;MYL6B	1.692
G2	ZBTB22;RGL2;PFDN6	1.416
G3	COL4A5;DPT;COL9A2;COL6A1;COL4A6	1.279
G4	NDST2;MOGS;GCNT1;GALNT3;B3GALT4	1.017
G5	NONO;KIAA0430;CELF2;RBPMS	0.848
G6	ZNF217;PRICKLE3;KIAA0913;ZMYM3;INO80B;CA9;WBP4	0.569
G7	PTPRN2;HPN;JAM3;SLC7A1	0.162
G8	ZIM2, PEG3;ZNF217;ZBTB22	0.082

Group is the group number. **Symbol** is the list of gene symbols that are present in each group. **Score** is the enrichment score for each group resulted from the functional analysis.

8.1.4 Chapter 5 - Results

Sample details of Alzheimer's Disease datasets.

Table 8.7: **Sample details of Alzheimer's Disease datasets.**

Accession	Region	Cell type	Ethnicity	Disease state	Sex	Age	ID
GSM119631	HIP	layer III neurons	Caucasian	normal	female	102 years	P1
GSM119643	MTG	layer III neurons	Caucasian	normal	female	102 years	P1
GSM119655	PC	layer III neurons	Caucasian	normal	female	102 years	P1
GSM119619	EC	layer III neurons	Caucasian	normal	female	102 years	P1
GSM119670	SFG	layer III neurons	Caucasian	normal	female	102 years	P1
GSM119615	EC	layer III neurons	Caucasian	normal	male	63 years	P2
GSM119632	HIP	layer III neurons	Caucasian	normal	male	63 years	P2
GSM119644	MTG	layer III neurons	Caucasian	normal	male	63 years	P2
GSM119656	PC	layer III neurons	Caucasian	normal	male	63 years	P2
GSM119671	SFG	layer III neurons	Caucasian	normal	male	63 years	P2
GSM119679	VCX	layer III neurons	Caucasian	normal	male	63 years	P2
GSM238824	MTG	layer III neurons	Caucasian	Alzheimer's Disease	male	68 years	P3
GSM238837	PC	pyramidal neuron	unknown	Alzheimer's Disease	male	68 years	P4
GSM238862	SFG	pyramidal neuron	unknown	Alzheimer's Disease	male	68 years	P4
GSM238854	SFG	pyramidal neuron	unknown	Alzheimer's Disease	male	68 years	P4
GSM238947	VCX	pyramidal neuron	unknown	Alzheimer's Disease	male	68 years	P4
GSM238943	VCX	pyramidal neuron	unknown	Alzheimer's Disease	male	68 years	P4
GSM238955	VCX	layer III neurons	Caucasian	Alzheimer's Disease	male	68 years	P4
GSM119626	EC	layer III neurons	Caucasian	normal	male	69 years	P5
GSM119639	HIP	layer III neurons	Caucasian	normal	male	69 years	P5

Sample details of Alzheimer's Disease datasets (continued)

Accession	Region	Cell type	Ethnicity	Disease state	Sex	Age	ID
GSM119651	MTG	layer III neurons	Caucasian	normal	male	69 years	P5
GSM119664	PC	layer III neurons	Caucasian	normal	male	69 years	P5
GSM119675	SFG	layer III neurons	Caucasian	normal	male	69 years	P5
GSM119687	VCX	layer III neurons	Caucasian	normal	male	69 years	P5
GSM238803	HIP	pyramidal neuron	unknown	Alzheimer's Disease	female	70.8 years	P6
GSM238838	PC	pyramidal neuron	unknown	Alzheimer's Disease	female	70.8 years	P6
GSM238856	SFG	pyramidal neuron	unknown	Alzheimer's Disease	female	70.8 years	P6
GSM238944	VCX	pyramidal neuron	unknown	Alzheimer's Disease	female	70.8 years	P6
GSM238808	HIP	layer III neurons	Caucasian	Alzheimer's Disease	male	72 years	P7
GSM238821	MTG	pyramidal neuron	unknown	Alzheimer's Disease	male	72 years	P8
GSM238810	MTG	pyramidal neuron	unknown	Alzheimer's Disease	male	72 years	P8
GSM238870	SFG	pyramidal neuron	unknown	Alzheimer's Disease	male	72 years	P8
GSM238844	SFG	pyramidal neuron	unknown	Alzheimer's Disease	male	72 years	P8
GSM238953	VCX	pyramidal neuron	unknown	Alzheimer's Disease	male	72 years	P8
GSM119638	HIP	layer III neurons	Caucasian	normal	female	73 years	P10
GSM119650	MTG	layer III neurons	Caucasian	normal	female	73 years	P10
GSM119663	PC	layer III neurons	Caucasian	normal	female	73 years	P10
GSM119686	VCX	layer III neurons	Caucasian	normal	female	73 years	P10
GSM119669	SFG	layer III neurons	Caucasian	normal	female	73 years	P10
GSM238826	PC	pyramidal neuron	unknown	Alzheimer's Disease	female	73 years	P9
GSM238799	HIP	pyramidal neuron	unknown	Alzheimer's Disease	female	73 years	P9
GSM238822	MTG	layer III neurons	Caucasian	Alzheimer's Disease	female	73 years	P11
GSM238842	SFG	pyramidal neuron	unknown	Alzheimer's Disease	female	73 years	P9
GSM238872	VCX	pyramidal neuron	unknown	Alzheimer's Disease	female	73 years	P9
GSM238868	SFG	pyramidal neuron	unknown	Alzheimer's Disease	male	74 years	P12
GSM238952	VCX	pyramidal neuron	unknown	Alzheimer's Disease	male	74 years	P12
GSM238802	HIP	pyramidal neuron	unknown	Alzheimer's Disease	male	75 years	P13
GSM238811	MTG	pyramidal neuron	unknown	Alzheimer's Disease	male	75 years	P13
GSM238813	MTG	pyramidal neuron	unknown	Alzheimer's Disease	male	75 years	P13
GSM238835	PC	pyramidal neuron	unknown	Alzheimer's Disease	male	75 years	P13
GSM238845	SFG	pyramidal neuron	unknown	Alzheimer's Disease	male	75 years	P13
GSM238847	SFG	pyramidal neuron	unknown	Alzheimer's Disease	male	75 years	P13
GSM238874	VCX	pyramidal neuron	unknown	Alzheimer's Disease	male	75 years	P13
GSM238877	VCX	pyramidal neuron	unknown	Alzheimer's Disease	male	75 years	P13
GSM119621	EC	layer III neurons	Caucasian	normal	male	76 years	P14
GSM119634	HIP	layer III neurons	Caucasian	normal	male	76 years	P14
GSM119681	VCX	layer III neurons	Caucasian	normal	male	76 years	P14
GSM119658	PC	layer III neurons	Caucasian	normal	male	76 years	P14
GSM119673	SFG	layer III neurons	Caucasian	normal	male	76 years	P14
GSM238805	HIP	pyramidal neuron	unknown	Alzheimer's Disease	female	77 years	P14
GSM238860	SFG	pyramidal neuron	unknown	Alzheimer's Disease	female	77 years	P14
GSM238794	EC	pyramidal neuron	unknown	Alzheimer's Disease	female	78 years	P15
GSM238801	HIP	pyramidal neuron	unknown	Alzheimer's Disease	male	78 years	P16
GSM238812	MTG	pyramidal neuron	unknown	Alzheimer's Disease	male	78 years	P16
GSM238834	PC	pyramidal neuron	unknown	Alzheimer's Disease	male	78 years	P16
GSM238846	SFG	pyramidal neuron	unknown	Alzheimer's Disease	male	78 years	P16
GSM238875	VCX	pyramidal neuron	unknown	Alzheimer's Disease	male	78 years	P16
GSM119627	EC	layer III neurons	Caucasian	normal	male	78 years	P17
GSM119640	HIP	layer III neurons	Caucasian	normal	male	78 years	P17
GSM119665	PC	layer III neurons	Caucasian	normal	male	78 years	P17
GSM119688	VCX	layer III neurons	Caucasian	normal	male	78 years	P17
GSM119652	MTG	layer III neurons	Caucasian	normal	male	78 years	P17
GSM238793	EC	pyramidal neuron	unknown	Alzheimer's Disease	female	79 years	P18
GSM238806	HIP	pyramidal neuron	unknown	Alzheimer's Disease	male	79 years	P19
GSM119620	EC	layer III neurons	Caucasian	normal	male	79 years	P20

Sample details of Alzheimer's Disease datasets (continued)

Accession	Region	Cell type	Ethnicity	Disease state	Sex	Age	ID
GSM119623	EC	layer III neurons	Caucasian	normal	male	79 years	P20
GSM119633	HIP	layer III neurons	Caucasian	normal	male	79 years	P20
GSM119636	HIP	layer III neurons	Caucasian	normal	male	79 years	P20
GSM119645	MTG	layer III neurons	Caucasian	normal	male	79 years	P20
GSM119647	MTG	layer III neurons	Caucasian	normal	male	79 years	P20
GSM119657	PC	layer III neurons	Caucasian	normal	male	79 years	P20
GSM119660	PC	layer III neurons	Caucasian	normal	male	79 years	P20
GSM119680	VCX	layer III neurons	Caucasian	normal	male	79 years	P20
GSM119683	VCX	layer III neurons	Caucasian	normal	male	79 years	P20
GSM119672	SFG	layer III neurons	Caucasian	normal	male	79 years	P20
GSM238818	MTG	pyramidal neuron	unknown	Alzheimer's Disease	male	79 years	P21
GSM238840	PC	pyramidal neuron	unknown	Alzheimer's Disease	male	79 years	P21
GSM238863	SFG	pyramidal neuron	unknown	Alzheimer's Disease	male	79 years	P21
GSM238948	VCX	pyramidal neuron	unknown	Alzheimer's Disease	male	79 years	P21
GSM238825	MTG	layer III neurons	Caucasian	Alzheimer's Disease	female	80 years	P22
GSM238963	VCX	layer III neurons	Caucasian	Alzheimer's Disease	female	80 years	P22
GSM238871	SFG	layer III neurons	Caucasian	Alzheimer's Disease	male	80 years	P23
GSM238798	EC	pyramidal neuron	unknown	Alzheimer's Disease	male	80 years	P24
GSM238867	SFG	pyramidal neuron	unknown	Alzheimer's Disease	female	80 years	P25
GSM119617	EC	layer III neurons	Caucasian	normal	male	80 years	P26
GSM119629	HIP	layer III neurons	Caucasian	normal	male	80 years	P26
GSM119630	HIP	layer III neurons	Caucasian	normal	male	80 years	P26
GSM119642	MTG	layer III neurons	Caucasian	normal	male	80 years	P26
GSM119654	PC	layer III neurons	Caucasian	normal	male	80 years	P26
GSM119678	VCX	layer III neurons	Caucasian	normal	male	80 years	P26
GSM119618	EC	layer III neurons	Caucasian	normal	male	80 years	P26
GSM238800	HIP	pyramidal neuron	unknown	Alzheimer's Disease	male	81 years	P27
GSM238809	MTG	pyramidal neuron	unknown	Alzheimer's Disease	male	81 years	P27
GSM238827	PC	pyramidal neuron	unknown	Alzheimer's Disease	male	81 years	P27
GSM238843	SFG	pyramidal neuron	unknown	Alzheimer's Disease	male	81 years	P27
GSM238873	VCX	pyramidal neuron	unknown	Alzheimer's Disease	male	81 years	P27
GSM238816	MTG	pyramidal neuron	unknown	Alzheimer's Disease	female	81.3 years	P28
GSM238945	VCX	pyramidal neuron	unknown	Alzheimer's Disease	female	81.3 years	P28
GSM238763	EC	pyramidal neuron	unknown	Alzheimer's Disease	female	82 years	P29
GSM238819	MTG	pyramidal neuron	unknown	Alzheimer's Disease	female	82 years	P29
GSM238864	SFG	pyramidal neuron	unknown	Alzheimer's Disease	female	82 years	P29
GSM238949	VCX	pyramidal neuron	unknown	Alzheimer's Disease	female	82 years	P29
GSM119625	EC	layer III neurons	Caucasian	normal	female	82 years	P30
GSM119649	MTG	layer III neurons	Caucasian	normal	female	82 years	P30
GSM119662	PC	layer III neurons	Caucasian	normal	female	82 years	P30
GSM119685	VCX	layer III neurons	Caucasian	normal	female	82 years	P30
GSM119668	SFG	layer III neurons	Caucasian	normal	female	82 years	P30
GSM238858	SFG	pyramidal neuron	unknown	Alzheimer's Disease	female	83 years	P31
GSM238861	SFG	pyramidal neuron	unknown	Alzheimer's Disease	female	83 years	P31
GSM119622	EC	layer III neurons	Caucasian	normal	male	83 years	P32
GSM119635	HIP	layer III neurons	Caucasian	normal	male	83 years	P32
GSM119646	MTG	layer III neurons	Caucasian	normal	male	83 years	P32
GSM119659	PC	layer III neurons	Caucasian	normal	male	83 years	P32
GSM119682	VCX	layer III neurons	Caucasian	normal	male	83 years	P32
GSM119674	SFG	layer III neurons	Caucasian	normal	male	83 years	P32
GSM238792	EC	pyramidal neuron	unknown	Alzheimer's Disease	male	84 years	P33
GSM119628	HIP	layer III neurons	Caucasian	normal	male	85 years	P34
GSM119616	EC	layer III neurons	Caucasian	normal	male	85 years	P34
GSM119641	MTG	layer III neurons	Caucasian	normal	male	85 years	P34
GSM119653	PC	layer III neurons	Caucasian	normal	male	85 years	P34

Sample details of Alzheimer's Disease datasets (continued)

Accession	Region	Cell type	Ethnicity	Disease state	Sex	Age	ID
GSM119677	VCX	layer III neurons	Caucasian	normal	male	85 years	P34
GSM238804	HIP	pyramidal neuron	unknown	Alzheimer's Disease	female	85 years	P35
GSM238817	MTG	pyramidal neuron	unknown	Alzheimer's Disease	female	85 years	P35
GSM238839	PC	pyramidal neuron	unknown	Alzheimer's Disease	female	85 years	P35
GSM238857	SFG	pyramidal neuron	unknown	Alzheimer's Disease	female	85 years	P35
GSM238946	VCX	pyramidal neuron	unknown	Alzheimer's Disease	female	85 years	P35
GSM238790	EC	pyramidal neuron	unknown	Alzheimer's Disease	female	86 years	P36
GSM238796	EC	pyramidal neuron	unknown	Alzheimer's Disease	male	86 years	P37
GSM238848	SFG	pyramidal neuron	unknown	Alzheimer's Disease	male	87 years	P38
GSM238941	VCX	pyramidal neuron	unknown	Alzheimer's Disease	male	87 years	P38
GSM238823	MTG	layer III neurons	Caucasian	Alzheimer's Disease	male	87 years	P39
GSM119624	EC	layer III neurons	Caucasian	normal	female	88 years	P40
GSM119637	HIP	layer III neurons	Caucasian	normal	female	88 years	P40
GSM119648	MTG	layer III neurons	Caucasian	normal	female	88 years	P40
GSM119684	VCX	layer III neurons	Caucasian	normal	female	88 years	P40
GSM119667	SFG	layer III neurons	Caucasian	normal	female	88 years	P40
GSM238807	HIP	pyramidal neuron	unknown	Alzheimer's Disease	male	88 years	P41
GSM238820	MTG	pyramidal neuron	unknown	Alzheimer's Disease	male	88 years	P41
GSM238841	PC	pyramidal neuron	unknown	Alzheimer's Disease	male	88 years	P41
GSM238865	SFG	pyramidal neuron	unknown	Alzheimer's Disease	male	88 years	P41
GSM238951	VCX	pyramidal neuron	unknown	Alzheimer's Disease	male	88 years	P41
GSM238795	EC	pyramidal neuron	unknown	Alzheimer's Disease	female	91 years	P42
GSM238791	EC	pyramidal neuron	unknown	Alzheimer's Disease	female	93 years	P43
GSM238815	MTG	pyramidal neuron	unknown	Alzheimer's Disease	female	95 years	P44
GSM238855	SFG	pyramidal neuron	unknown	Alzheimer's Disease	female	95 years	P44
GSM238851	SFG	pyramidal neuron	unknown	Alzheimer's Disease	female	95 years	P44
GSM238942	VCX	pyramidal neuron	unknown	Alzheimer's Disease	female	95 years	P44

Accession is the GEO accession number for each sample. **Region** is the name of brain regions considered for the study. **Cell type** is the type of cell for each sample. **Ethnicity, Disease state, Sex, Age and ID** are the details related with the person where the particular sample is taken from.

The details of Coloured (α, β)-k Feature Set problem result with run time, coverage,etc.

Table 8.8: The details of the Coloured (α, β)-k Feature Set problem result.

Features: 3120, Cases:126, Alfa Nodes:754, Beta Nodes:814
Maximum Alfa: 2725, Maximum Beta: 2789
Minimum Alfa: 396, Minimum Beta: 865
NA value: -1, Classes: 2, 259 Beta target:true, 167 Beta target:true, Colours: 5
SFG-32,PC-21,HIP-23,MTG-28,EC-22
Beta targets: 2
Has Targets: false
Has Case Weights: true
Has Case Adjacency: true
Alfa Adjacency: true
Beta Adjacency: true
Negate Beta Adjacency: false
Has Feature Weights: true
Working Memory : 256.000000% (WorkMem=256.000000)
Presolve time = 0.86 sec. (530.35 ticks)
Probing time = 0.02 sec. (8.93 ticks)
Presolve time = 0.46 sec. (287.74 ticks)
Probing time = 0.02 sec. (8.93 ticks)
Parallel mode: deterministic, using up to 32 threads.
Root relaxation solution time = 0.17 sec. (350.15 ticks)
Root node processing (before b&c):
Real time = 3.42 sec. (2400.86 ticks)
Parallel b&c, 32 threads:
Real time = 0.31 sec. (22.50 ticks)
Sync time (average) = 0.03 sec.
Wait time (average) = 0.00 sec.
Total (root+branch&cut) = 3.73 sec. (2423.36 ticks)
Solution for model: Max Sum alfa beta cover
Solution status = Optimal
Solution value = 903802
Solution time = 14.95
Condition: Alfa: 396 Beta: 300 K: 825 Option:gap:0
**Solution: Optimal Cover: 903802
————————————

The list of probes that are common in the result of from Coloured (α, β)-k Feature Set and Generalised (α, β)-k Feature Set problem approach.

Table 8.9: **The list of probes resulted from Coloured (α, β)-k Feature Set and Generalised (α, β)-k Feature Set problem approach**

Probe ID	Gene	Symbol
1316_at	thyroid hormone receptor, alpha	THRA
1552507_at	potassium channel, voltage gated subfamily E regulatory beta subunit 4	KCNE4
1553186_x_at	RAS and EF-hand domain containing	RASEF
1553693_s_at	carbonyl reductase 4	CBR4
1554593_s_at	solute carrier family 1 (high affinity aspartate/glutamate transporter), member 6	SLC1A6
1555495_a_at	CWC27 spliceosome-associated protein homolog (S. cerevisiae)	CWC27
1556690_s_at	Data not found	
1557155_a_at	Data not found	
1557293_at	long intergenic non-protein coding RNA 969	LINC00969
1557895_at	FLJ35934	FLJ35934
1558279_a_at	3-ketodihydrosphingosine reductase	KDSR
1558678_s_at	metastasis associated lung adenocarcinoma transcript 1 (non-protein coding)	MALAT1
1558695_at	Data not found	
1558792_x_at	adaptor-related protein complex 2, alpha 1 subunit	AP2A1
1558831_x_at	Data not found	
1559060_a_at	folliculin interacting protein 1	FNIP1
1559391_s_at	Data not found	
1559949_at	Data not found	
1560116_a_at	neural precursor cell expressed, developmentally down-regulated 1	NEDD1
1560689_s_at	v-akt murine thymoma viral oncogene homolog 2	AKT2
1560741_at	small nuclear ribonucleoprotein polypeptide N	SNRPN
1563781_at	hypothetical protein LOC285949	LOC285949
1563881_at	Data not found	
1565620_at	ArfGAP with GTPase domain, ankyrin repeat and PH domain 4	AGAP4
1566480_x_at	chromosome 17 open reading frame 104	C17orf104
1568603_at	Ca++-dependent secretion activator	CADPS
1568612_at	gamma-aminobutyric acid (GABA) A receptor, gamma 2	GABRG2
1568763_s_at	In multiple Geneids	
1568877_a_at	acyl-CoA binding domain containing 5	ACBD5
1569200_at	Data not found	
1569302_at	centrosomal protein 295kDa	CEP295
1569482_at	Data not found	
1569661_at	neuronal PAS domain protein 3	NPAS3
1570210_x_at	protein phosphatase 6, regulatory subunit 2	PPP6R2
200039_s_at	proteasome (prosome, macropain) subunit, beta type, 2	PSMB2
200042_at	RNA 2',3'-cyclic phosphate and 5'-OH ligase	RTCB
200047_s_at	YY1 transcription factor	YY1
200053_at	sperm associated antigen 7	SPAG7
200072_s_at	heterogeneous nuclear ribonucleoprotein M	HNRNPM
200085_s_at	transcription elongation factor B (SIII), polypeptide 2 (18kDa, elongin B)	TCEB2
200098_s_at	anaphase promoting complex subunit 5	ANAPC5
200625_s_at	CAP, adenylate cyclase-associated protein 1 (yeast)	CAP1
200639_s_at	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta	YWHAZ

The list of probes resulted from Coloured (α, β)-*k* Feature Set and Generalised (α, β)-*k* Feature Set problem approach (continued)

Probe ID	Gene	Symbol
200640_at	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta	YWHAZ
200685_at	serine/arginine-rich splicing factor 11	SRSF11
200693_at	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, theta	YWHAQ
200708_at	glutamic-oxaloacetic transaminase 2, mitochondrial	GOT2
200732_s_at	protein tyrosine phosphatase type IVA, member 1	PTP4A1
200754_x_at	serine/arginine-rich splicing factor 2	SRSF2
200771_at	laminin, gamma 1 (formerly LAMB2)	LAMC1
200775_s_at	heterogeneous nuclear ribonucleoprotein K	HNRNPK
200822_x_at	triosephosphate isomerase 1	TPI1
200954_at	ATPase, H+ transporting, lysosomal 16kDa, V0 subunit c	ATP6V0C
200976_s_at	Tax1 (human T-cell leukemia virus type I) binding protein 1	TA
200980_s_at	pyruvate dehydrogenase (lipoamide) alpha 1	PDHA1
201057_s_at	golgin B1	GOLGB1
201065_s_at	In multiple Geneids	
201067_at	proteasome (prosome, macropain) 26S subunit, ATPase, 2	PSMC2
201083_s_at	BCL2-associated transcription factor 1	BCLAF1
201145_at	HCLS1 associated protein	
201146_at	nuclear factor, erythroid 2-like 2	NFE2L2
201182_s_at	chromodomain helicase DNA binding protein 4	CHD4
201217_x_at	ribosomal protein L3	RPL3
201290_at	SEC11 homolog A (S. cerevisiae)	SEC11A
201305_x_at	acidic (leucine-rich) nuclear phosphoprotein 32 family, member B	ANP32B
201324_at	epithelial membrane protein 1	EMP1
201371_s_at	cullin 3	CUL3
201400_at	proteasome (prosome, macropain) subunit, beta type, 3	PSMB3
201410_at	pleckstrin homology domain containing, family B (evectins) member 2	PLEKHB2
201415_at	glutathione synthetase	GSS
201434_at	tetratricopeptide repeat domain 1	TTC1
201441_at	cytochrome c oxidase subunit VIb polypeptide 1 (ubiquitous)	CO
201500_s_at	protein phosphatase 1, regulatory (inhibitor) subunit 11	PPP1R11
201509_at	isocitrate dehydrogenase 3 (NAD+) beta	IDH3B
201527_at	ATPase, H+ transporting, lysosomal 14kDa, V1 subunit F	ATP6V1F
201548_s_at	lysine (K)-specific demethylase 5B	KDM5B
201570_at	SAMM50 sorting and assembly machinery component	SAMM50
201586_s_at	splicing factor proline/glutamine-rich	SFPQ
201622_at	staphylococcal nuclease and tudor domain containing 1	SND1
201636_at	fragile	
201704_at	ectonucleoside triphosphate diphosphohydrolase 6 (putative)	ENTPD6
201709_s_at	nipsnap homolog 1 (C. elegans)	NIPSNAP1
201810_s_at	SH3-domain binding protein 5 (BTK-associated)	SH3BP5
201828_x_at	family with sequence similarity 127, member A	FAM127A
201836_s_at	suppressor of Ty 7 (S. cerevisiae)-like	SUPT7L
201991_s_at	kinesin family member 5B	KIF5B
202025_x_at	acetyl-CoA acyltransferase 1	ACAA1
202070_s_at	isocitrate dehydrogenase 3 (NAD+) alpha	IDH3A
202120_x_at	adaptor-related protein complex 2, sigma 1 subunit	AP2S1
202121_s_at	charged multivesicular body protein 2A	CHMP2A
202135_s_at	ARP1 actin-related protein 1 homolog B, centractin beta (yeast)	ACTR1B
202138_x_at	aminoacyl tRNA synthetase complex-interacting multifunctional protein 2	AIMP2
202144_s_at	adenylosuccinate lyase	ADSL
202160_at	CREB binding protein	CREBBP

The list of probes resulted from Coloured (α, β)-*k* Feature Set and Generalised (α, β)-*k* Feature Set problem approach (continued)

Probe ID	Gene	Symbol
202178_at	protein kinase C, zeta	PRKCZ
202201_at	biliverdin reductase B	BLVRB
202330_s_at	uracil-DNA glycosylase	UNG
202360_at	mastermind-like 1 (Drosophila)	MAML1
202505_at	small nuclear ribonucleoprotein polypeptide B	SNRPB2
202534_x_at	dihydrofolate reductase	DHFR
202551_s_at	cysteine rich transmembrane BMP regulator 1 (chordin-like)	CRIM1
202594_at	leptin receptor overlapping transcript-like 1	LEPROTL1
202712_s_at	In multiple Geneids	
202717_s_at	cell division cycle 16	CDC16
202820_at	aryl hydrocarbon receptor	AHR
202829_s_at	vesicle-associated membrane protein 7	VAMP7
202858_at	U2 small nuclear RNA auxiliary factor 1	U2AF1
202868_s_at	processing of precursor 4, ribonuclease P/MRP subunit (S. cerevisiae)	POP4
202927_at	peptidylprolyl cis/trans isomerase, NIMA-interacting 1	PIN1
202961_s_at	ATP synthase, H+ transporting, mitochondrial Fo complex, subunit F2	ATP5J2
202974_at	membrane protein, palmitoylated 1, 55kDa	MPP1
203031_s_at	uroporphyrinogen III synthase	UROS
203068_at	kelch-like family member 21	KLHL21
203082_at	BMS1 ribosome biogenesis factor	BMS1
203113_s_at	eukaryotic translation elongation factor 1 delta (guanine nucleotide exchange protein)	EEF1D
203122_at	trafficking protein particle complex 12	TRAPPC12
203132_at	retinoblastoma 1	RB1
203140_at	B-cell CLL/lymphoma 6	BCL6
203146_s_at	gamma-aminobutyric acid (GABA) B receptor, 1	GABBR1
203297_s_at	jumonji, AT rich interactive domain 2	JARID2
203431_s_at	Rho GTPase activating protein 32	ARHGAP32
203466_at	MpV17 mitochondrial inner membrane protein	MPV17
203485_at	reticulon 1	RTN1
203509_at	sortilin-related receptor, L(DLR class) A repeats containing	SORL1
203549_s_at	lipoprotein lipase	LPL
203606_at	NADH dehydrogenase (ubiquinone) Fe-S protein 6, 13kDa (NADH-coenzyme Q reductase)	NDUFS6
203752_s_at	jun D proto-oncogene	JUND
203773_x_at	biliverdin reductase A	BLVRA
203814_s_at	NAD(P)H dehydrogenase, quinone 2	NQO2
203849_s_at	kinesin family member 1A	KIF1A
203909_at	solute carrier family 9, subfamily A (NHE6, cation proton antiporter 6), member 6	SLC9A6
203911_at	RAP1 GTPase activating protein	RAP1GAP
203956_at	MORC family CW-type zinc finger 2	MORC2
204229_at	solute carrier family 17 (vesicular glutamate transporter), member 7	SLC17A7
204266_s_at	choline kinase alpha	CHKA
204270_at	SKI proto-oncogene	SKI
204321_at	neogenin 1	NEO1
204466_s_at	synuclein, alpha (non A4 component of amyloid precursor)	SNCA
204481_at	bromodomain and PHD finger containing, 1	BRPF1
204517_at	peptidylprolyl isomerase C (cyclophilin C)	PPIC
204528_s_at	nucleosome assembly protein 1-like 1	NAP1L1
204552_at	inositol polyphosphate-4-phosphatase, type I, 107kDa	INPP4A
204559_s_at	LSM7 homolog, U6 small nuclear RNA associated (S. cerevisiae)	LSM7
204584_at	L1 cell adhesion molecule	L1CAM
204622_x_at	nuclear receptor subfamily 4, group A, member 2	NR4A2

The list of probes resulted from Coloured (α, β)-*k* Feature Set and Generalised (α, β)-*k* Feature Set problem approach (continued)

Probe ID	Gene	Symbol
204663_at	malic enzyme 3, NADP(+)-dependent, mitochondrial	ME3
204680_s_at	Rap guanine nucleotide exchange factor (GEF) 5	RAPGEF5
204718_at	EPH receptor B6	EPHB6
204720_s_at	DnaJ (Hsp40) homolog, subfamily C, member 6	DNAJC6
204731_at	transforming growth factor, beta receptor III	TGFBR3
204786_s_at	interferon (alpha, beta and omega) receptor 2	IFNAR2
204957_at	origin recognition complex, subunit 5	ORC5
204977_at	DEAD (Asp-Glu-Ala-Asp) box polypeptide 10	DD
205117_at	fibroblast growth factor 1 (acidic)	FGF1
205257_s_at	amphiphysin	AMPH
205353_s_at	phosphatidylethanolamine binding protein 1	PEBP1
205514_at	zinc finger protein 415	ZNF415
205559_s_at	proprotein convertase subtilisin/kexin type 5	PCSK5
205570_at	phosphatidylinositol-5-phosphate 4-kinase, type II, alpha	PIP4K2A
205594_at	zinc finger protein 652	ZNF652
205751_at	SH3-domain GRB2-like 2	SH3GL2
205787_x_at	zinc finger CCCH-type containing 11A	ZC3H11A
205816_at	integrin, beta 8	ITGB8
205967_at	histone cluster 1, H4c	HIST1H4C
206122_at	SRY (sex determining region Y)-box 15	SO
206140_at	LIM homeobox 2	LH
206273_at	slowmo homolog 1 (Drosophila)	SLMO1
206307_s_at	forkhead box D1	FO
206547_s_at	protein phosphatase, EF-hand calcium binding domain 1	PPEF1
206562_s_at	casein kinase 1, alpha 1	CSNK1A1
206652_at	zinc finger, MYM-type 5	ZMYM5
207081_s_at	phosphatidylinositol 4-kinase, catalytic, alpha	PI4KA
207084_at	POU class 3 homeobox 2	POU3F2
207132_x_at	prefoldin subunit 5	PFDN5
207598_x_at		
207789_s_at	dipeptidyl-peptidase 6	DPP6
207966_s_at	golgi glycoprotein 1	GLG1
208549_x_at	prothymosin, alpha	PTMA
208611_s_at	spectrin, alpha, non-erythrocytic 1	SPTAN1
208704_x_at	amyloid beta (A4) precursor-like protein 2	APLP2
208710_s_at	adaptor-related protein complex 3, delta 1 subunit	AP3D1
208732_at	RAB2A, member RAS oncogene family	RAB2A
208835_s_at	LUC7-like 3 (S. cerevisiae)	LUC7L3
208859_s_at	alpha thalassemia/mental retardation syndrome	
208936_x_at	lectin, galactoside-binding, soluble, 8	LGALS8
208942_s_at	SEC62 homolog (S. cerevisiae)	SEC62
208955_at	deoxyuridine triphosphatase	DUT
208969_at	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 9, 39kDa	NDUFA9
208972_s_at	ATP synthase, H+ transporting, mitochondrial Fo complex, subunit C1 (subunit 9)	ATP5G1
208996_s_at	polymerase (RNA) II (DNA directed) polypeptide C, 33kDa	POLR2C
209001_s_at	anaphase promoting complex subunit 13	ANAPC13
209023_s_at	stromal antigen 2	STAG2
209066_x_at	ubiquinol-cytochrome c reductase binding protein	UQCRB
209079_x_at	In multiple Geneids	
209163_at	cytochrome b561	CYB561
209164_s_at	cytochrome b561	CYB561
209169_at	glycoprotein M6B	GPM6B
209177_at	NADH dehydrogenase (ubiquinone) complex I, assembly factor 3	NDUFAF3

The list of probes resulted from Coloured (α, β)-*k* Feature Set and Generalised (α, β)-*k* Feature Set problem approach (continued)

Probe ID	Gene	Symbol
209225_x_at	transportin 1	TNPO1
209250_at	delta(4)-desaturase, sphingolipid 1	DEGS1
209251_x_at	tubulin, alpha 1c	TUBA1C
209258_s_at	structural maintenance of chromosomes 3	SMC3
209268_at	vacuolar protein sorting 45 homolog (S. cerevisiae)	VPS45
209289_at	nuclear factor I/B	NFIB
209316_s_at	HBS1-like translational GTPase	HBS1L
209343_at	EF-hand domain family, member D1	EFHD1
209409_at	growth factor receptor-bound protein 10	GRB10
209440_at	phosphoribosyl pyrophosphate synthetase 1	PRPS1
209534_x_at	A kinase (PRKA) anchor protein 13	AKAP13
209553_at	vacuolar protein sorting 8 homolog (S. cerevisiae)	VPS8
209558_s_at	In multiple Geneids	
209583_s_at	CD200 molecule	CD200
209586_s_at	prune exopolyphosphatase	PRUNE
209609_s_at	mitochondrial ribosomal protein L9	MRPL9
209686_at	S100 calcium binding protein B	S100B
209715_at	chromobox homolog 5	CB
209751_s_at	In multiple Geneids	
209991_x_at	gamma-aminobutyric acid (GABA) B receptor, 2	GABBR2
210111_s_at	PNAS-119	LOC100287552
210268_at	nuclear transcription factor,	
210679_x_at	Data not found	
210686_x_at	solute carrier family 25 (mitochondrial carrier), member 16	SLC25A16
210701_at	craniofacial development protein 1	CFDP1
210825_s_at	phosphatidylethanolamine binding protein 1	PEBP1
210840_s_at	IQ motif containing GTPase activating protein 1	IQGAP1
210908_s_at	prefoldin subunit 5	PFDN5
210976_s_at	phosphofructokinase, muscle	PFKM
211386_at	uncharacterized protein MGC12488	MGC12488
211464_x_at	caspase 6, apoptosis-related cysteine peptidase	CASP6
211616_s_at	5-hydroxytryptamine (serotonin) receptor 2A, G protein-coupled	HTR2A
211752_s_at	NADH dehydrogenase (ubiquinone) Fe-S protein 7, 20kDa (NADH-coenzyme Q reductase)	NDUFS7
211779_x_at	adaptor-related protein complex 2, alpha 2 subunit	AP2A2
211921_x_at	prothymosin, alpha	PTMA
211930_at	heterogeneous nuclear ribonucleoprotein A3	HNRNPA3
211964_at	collagen, type IV, alpha 2	COL4A2
211993_at	WNK lysine deficient protein kinase 1	WNK1
212088_at	peptidase (mitochondrial processing) alpha	PMPCA
212095_s_at	microtubule associated tumor suppressor 1	MTUS1
212104_s_at	RNA binding protein, fox-1 homolog (C. elegans) 2	RBFO
212114_at	ataxin 7-like 3B	AT
212177_at	PNN-interacting serine/arginine-rich protein	PNISR
212178_s_at	In multiple Geneids	
212207_at	mediator complex subunit 13-like	MED13L
212209_at	mediator complex subunit 13-like	MED13L
212228_s_at	coenzyme Q9	COQ9
212252_at	calcium/calmodulin-dependent protein kinase kinase 2, beta	CAMKK2
212265_at	QKI, KH domain containing, RNA binding	QKI
212296_at	proteasome (prosome, macropain) 26S subunit, non-ATPase, 14	PSMD14
212322_at	sphingosine-1-phosphate lyase 1	SGPL1
212348_s_at	lysine (K)-specific demethylase 1A	KDM1A
212372_at	myosin, heavy chain 10, non-muscle	MYH10

The list of probes resulted from Coloured (α, β)-k Feature Set and Generalised (α, β)-k Feature Set problem approach (continued)

Probe ID	Gene	Symbol
212384_at	DEAD (Asp-Glu-Ala-Asp) box polypeptide 39B	DD
212432_at	GrpE-like 1, mitochondrial (E. coli)	GRPEL1
212450_at	SECIS binding protein 2-like	SECISBP2L
212451_at	SECIS binding protein 2-like	SECISBP2L
212462_at	K(lysine) acetyltransferase 6B	KAT6B
212468_at	sperm associated antigen 9	SPAG9
212483_at	Nipped-B homolog (Drosophila)	NIPBL
212491_s_at	DnaJ (Hsp40) homolog, subfamily C, member 8	DNAJC8
212501_at	CCAAT/enhancer binding protein (C/EBP), beta	CEBPB
212506_at	phosphatidylinositol binding clathrin assembly protein	PICALM
212581_x_at	glyceraldehyde-3-phosphate dehydrogenase	GAPDH
212645_x_at	brain and reproductive organ-expressed (TNFRSF1A modulator)	BRE
212652_s_at	sorting nexin 4	SN
212673_at	methionyl aminopeptidase 1	METAP1
212832_s_at	cytoskeleton associated protein 5	CKAP5
212896_at	superkiller viralicidic activity 2-like 2 (S. cerevisiae)	SKIV2L2
212992_at	AHNAK nucleoprotein 2	AHNAK2
212995_x_at	mitotic spindle organizing protein 2B	MZT2B
213015_at	bobby sox homolog (Drosophila)	BB
213039_at	Rho/Rac guanine nucleotide exchange factor (GEF) 18	ARHGEF18
213079_at	TSR2, 20S rRNA accumulation, homolog (S. cerevisiae)	TSR2
213089_at	uncharacterized LOC100272216	LOC100272216
213093_at	protein kinase C, alpha	PRKCA
213195_at	LYR motif containing 9	LYRM9
213298_at	nuclear factor I/C (CCAAT-binding transcription factor)	NFIC
213421_x_at	protease, serine, 3	PRSS3
213453_x_at	glyceraldehyde-3-phosphate dehydrogenase	GAPDH
213457_at	malignant fibrous histiocytoma amplified sequence 1	MFHAS1
213517_at	poly(rC) binding protein 2	PCBP2
213535_s_at	ubiquitin-conjugating enzyme E2I	UBE2I
213545_x_at	sorting nexin 3	SN
213693_s_at	mucin 1, cell surface associated	MUC1
213710_s_at	calmodulin 1 (phosphorylase kinase, delta)	CALM1
213744_at	attractin-like 1	ATRNL1
213808_at	Data not found	
214043_at	protein tyrosine phosphatase, receptor type, D	PTPRD
214246_x_at	misshapen-like kinase 1	MINK1
214375_at	In multiple Geneids	
214394_x_at	eukaryotic translation elongation factor 1 delta (guanine nucleotide exchange protein)	EEF1D
214436_at	F-box and leucine-rich repeat protein 2	FB
214499_s_at	BCL2-associated transcription factor 1	BCLAF1
214594_x_at	ATPase, aminophospholipid transporter, class I, type 8B, member 1	ATP8B1
214623_at	F-box and WD repeat domain containing 4 pseudogene 1	FB
214707_x_at	Alstrom syndrome protein 1	ALMS1
214743_at	cut-like homeobox 1	CU
214799_at	neurofascin	NFASC
214823_at	zinc finger protein 204, pseudogene	ZNF204P
214850_at	glucuronidase, beta pseudogene	LOC100170939
214934_at	ATPase, class II, type 9B	ATP9B
215021_s_at	neurexin 3	NR
215091_s_at	general transcription factor IIIA	GTF3A
215167_at	mediator complex subunit 14	MED14

The list of probes resulted from Coloured (α, β)-*k* Feature Set and Generalised (α, β)-*k* Feature Set problem approach (continued)

Probe ID	Gene	Symbol
215280_s_at	protein tyrosine phosphatase, receptor type, f polypeptide (PTPRF), interacting protein (liprin), alpha 3	PPFIA3
215306_at	Data not found	
215383_x_at	spastic paraplegia 21 (autosomal recessive, Mast syndrome)	SPG21
215385_at	Data not found	
215493_x_at	butyrophilin, subfamily 2, member A1	BTN2A1
215504_x_at	Data not found	
215514_at	Data not found	
215543_s_at	like-glycosyltransferase	LARGE
215553_x_at	Data not found	
215600_x_at	F-box and WD repeat domain containing 12	FB
215698_at	lysine (K)-specific demethylase 5A	KDM5A
215889_at	SKI-like proto-oncogene	SKIL
215963_x_at	In multiple Geneids	
216101_at	Data not found	
216308_x_at	glyoxylate reductase/hydroxypyruvate reductase	GRHPR
216524_x_at	Data not found	
216527_at	Data not found	
216550_x_at	ankyrin repeat domain 12	ANKRD12
216745_x_at	Data not found	
217028_at	chemokine (C-	
217077_s_at	gamma-aminobutyric acid (GABA) B receptor, 2	GABBR2
217250_s_at	chromodomain helicase DNA binding protein 5	CHD5
217398_x_at	glyceraldehyde-3-phosphate dehydrogenase	GAPDH
217482_at	Data not found	
217713_x_at	Data not found	
217726_at	coatomer protein complex, subunit zeta 1	COPZ1
217761_at	acireductone dioxygenase 1	ADI1
217819_at	golgin A7	GOLGA7
217860_at	In multiple Geneids	
217871_s_at	macrophage migration inhibitory factor (glycosylation-inhibiting factor)	MIF
217904_s_at	beta-site APP-cleaving enzyme 1	BACE1
217927_at	signal peptidase complex subunit 1 homolog (S. cerevisiae)	SPCS1
217937_s_at	histone deacetylase 7	HDAC7
217939_s_at	aftiphilin	AFTPH
217946_s_at	SUMO1 activating enzyme subunit 1	SAE1
217969_at	vacuolar protein sorting 51 homolog (S. cerevisiae)	VPS51
218026_at	cytochrome c oxidase assembly factor 3	COA3
218048_at	COMM domain containing 3	COMMD3
218247_s_at	mex-3 RNA binding family member C	ME
218271_s_at	presenilin associated, rhomboid-like	PARL
218298_s_at	chromosome 14 open reading frame 159	C14orf159
218302_at	presenilin enhancer gamma secretase subunit	PSENEN
218330_s_at	neuron navigator 2	NAV2
218381_s_at	U2 small nuclear RNA auxiliary factor 2	U2AF2
218520_at	TANK-binding kinase 1	TBK1
218522_s_at	microtubule-associated protein 1S	MAP1S
218543_s_at	poly (ADP-ribose) polymerase family, member 12	PARP12
218843_at	fibronectin type III domain containing 4	FNDC4
218865_at	mitochondrial amidoxime reducing component 1	MARCH1
218953_s_at	prenylcysteine oxidase 1 like	PCYO
219019_at	p53-induced death domain protein 1	PIDD1
219028_at	homeodomain interacting protein kinase 2	HIPK2
219032_x_at	opsin 3	OPN3

The list of probes resulted from Coloured (α, β)-k Feature Set and Generalised (α, β)-k Feature Set problem approach (continued)

Probe ID	Gene	Symbol
219204_s_at	serine racemase	SRR
219387_at	coiled-coil domain containing 88A	CCDC88A
219389_at	sushi domain containing 4	SUSD4
219670_at	BEN domain containing 5	BEND5
219671_at	hippocalcin like 4	HPCAL4
219709_x_at	family with sequence similarity 173, member A	FAM173A
219961_s_at	kizuna centrosomal protein	KIZ
220071_x_at	HAUS augmin-like complex, subunit 2	HAUS2
220072_at	centrosome and spindle pole associated protein 1	CSPP1
220081_x_at	hydroxysteroid (17-beta) dehydrogenase 7	HSD17B7
220136_s_at	crystallin, beta A2	CRYBA2
220269_at	zinc finger, B-box domain containing	ZBB
220295_x_at	DEP domain containing 1	DEPDC1
220316_at	neuronal PAS domain protein 3	NPAS3
220609_at	SUMO-interacting motifs containing 1 pseudogene	LOC202181
220615_s_at	fatty acyl CoA reductase 2	FAR2
220942_x_at	family with sequence similarity 162, member A	FAM162A
220966_x_at	actin related protein 2/3 complex, subunit 5-like	ARPC5L
221476_s_at	ribosomal protein L15	RPL15
221486_at	endosulfine alpha	ENSA
221488_s_at	cutA divalent cation tolerance homolog (E. coli)	CUTA
221497_x_at	egl-9 family hypoxia-inducible factor 1	EGLN1
221506_s_at	transportin 2	TNPO2
221772_s_at	protein phosphatase 2, regulatory subunit B, delta	PPP2R2D
221853_s_at	In multiple Geneids	
221864_at	ORAI calcium release-activated calcium modulator 3	ORAI3
221877_at	immunity-related GTPase family, Q	IRGQ
221886_at	DENN/MADD domain containing 2A	DENND2A
221942_s_at	guanylate cyclase 1, soluble, alpha 3	GUCY1A3
221952_x_at	tRNA methyltransferase 5	TRMT5
222024_s_at	A kinase (PRKA) anchor protein 13	AKAP13
222043_at	clusterin	CLU
222047_s_at	serrate, RNA effector molecule	SRRT
222154_s_at	spermatogenesis associated, serine-rich 2-like	SPATS2L
222160_at	A kinase (PRKA) anchor protein 8-like	AKAP8L
222284_at	Data not found	
222364_at	solute carrier family 44 (choline transporter), member 1	SLC44A1
222380_s_at	programmed cell death 6	PDCD6
222395_s_at	ubiquitin-conjugating enzyme E2Z	UBE2Z
222428_s_at	leucyl-tRNA synthetase	LARS
222439_s_at	thyroid hormone receptor associated protein 3	THRAP3
222457_s_at	LIM domain and actin binding 1	LIMA1
222513_s_at	sorbin and SH3 domain containing 1	SORBS1
222610_s_at	S100P binding protein	S100PBP
222651_s_at	trichorhinophalangeal syndrome I	TRPS1
222745_s_at	katanin p80 subunit B-like 1	KATNBL1
222809_x_at	coiled-coil domain containing 85C	CCDC85C
222982_x_at	solute carrier family 38, member 2	SLC38A2
223024_at	adaptor-related protein complex 1, mu 1 subunit	AP1M1
223035_s_at	phenylalanyl-tRNA synthetase, beta subunit	FARSB
223037_at	PDZ domain containing 11	PDZD11
223134_at	bobby sox homolog (Drosophila)	BB
223185_s_at	basic helix-loop-helix family, member e41	BHLHE41
223319_at	gephyrin	GPHN

The list of probes resulted from Coloured (α, β)-*k* Feature Set and Generalised (α, β)-*k* Feature Set problem approach (continued)

Probe ID	Gene	Symbol
223380_s_at	large tumor suppressor kinase 2	LATS2
223480_s_at	mitochondrial ribosomal protein L47	MRPL47
223519_at	sterile alpha motif and leucine zipper containing kinase AZK	ZAK
223673_at	regulatory factor	
223679_at	catenin (cadherin-associated protein), beta 1, 88kDa	CTNNB1
224151_s_at	adenylate kinase 3	AK3
224227_s_at	B double prime 1, subunit of RNA polymerase III transcription initiation factor IIIB	BDP1
224366_s_at	RALBP1 associated Eps domain containing 1	REPS1
224369_s_at	F-box protein 38	FB
224414_s_at	caspase recruitment domain family, member 6	CARD6
224568_x_at	metastasis associated lung adenocarcinoma transcript 1 (non-protein coding)	MALAT1
224598_at	mannosyl (alpha-1,3-)-glycoprotein beta-1,4-N-acetylglucosaminyltransferase, isozyme B	MGAT4B
224613_s_at	DnaJ (Hsp40) homolog, subfamily C, member 5	DNAJC5
224628_at	endoplasmic reticulum lectin 1	ERLEC1
224666_at	non-SMC element 1 homolog (S. cerevisiae)	NSMCE1
224689_at	mannosidase, beta A, lysosomal-like	MANBAL
224737_x_at	cell division cycle and apoptosis regulator 1	CCAR1
224739_at	Pim-3 proto-oncogene, serine/threonine kinase	PIM3
224774_s_at	neuron navigator 1	NAV1
224862_at	guanine nucleotide binding protein (G protein), q polypeptide	GNAQ
224878_at	ubiquitin family domain containing 1	UBFD1
224972_at	reactive oxygen species modulator 1	ROMO1
224999_at	Data not found	
225003_at	transmembrane protein 205	TMEM205
225050_at	zinc finger protein 512	ZNF512
225055_at	leucine rich repeat containing 37, member A16, pseudogene	LRRC37A16P
225092_at	rabaptin, RAB GTPase binding effector protein 1	RABEP1
225172_at	Crm, cramped-like (Drosophila)	CRAMP1L
225219_at	SMAD family member 5	SMAD5
225223_at	SMAD family member 5	SMAD5
225234_at	Cbl proto-oncogene, E3 ubiquitin protein ligase	CBL
225239_at	Data not found	
225240_s_at	musashi RNA-binding protein 2	MSI2
225298_at	paroxysmal nonkinesigenic dyskinesia	PNKD
225356_at	Data not found	
225463_x_at	G protein-coupled receptor 89A	GPR89A
225484_at	centrosomal protein 41kDa	CEP41
225493_at	uncharacterized LOC144438	LOC144438
225496_s_at	synaptotagmin-like 2	SYTL2
225501_at	PHD finger protein 6	PHF6
225516_at	solute carrier family 7 (cationic amino acid transporter, y+ system), member 2	SLC7A2
225571_at	leukemia inhibitory factor receptor alpha	LIFR
225636_at	signal transducer and activator of transcription 2, 113kDa	STAT2
225649_s_at	serine/threonine kinase 35	STK35
225655_at	ubiquitin-like with PHD and ring finger domains 1	UHRF1
225750_at	Data not found	
225864_at	family with sequence similarity 84, member B	FAM84B
225870_s_at	trafficking protein particle complex 5	TRAPPC5
225917_at	Data not found	
225923_at	Data not found	

The list of probes resulted from Coloured (α, β)-*k* Feature Set and Generalised (α, β)-*k* Feature Set problem approach (continued)

Probe ID	Gene	Symbol
225961_at	kelch-like family member 42	KLHL42
226037_s_at	TAF9B RNA polymerase II, TATA box binding protein (TBP)-associated factor, 31kDa	TAF9B
226086_at	synaptotagmin	
226101_at	protein kinase C, epsilon	PRKCE
226112_at	sarcoglycan, beta (43kDa dystrophin-associated glycoprotein)	SGCB
226143_at	retinoic acid induced 1	RAI1
226169_at	SET binding factor 2	SBF2
226195_at	intraflagellar transport 43	IFT43
226223_at	Data not found	
226447_at	ash1 (absent, small, or homeotic)-like (Drosophila)	ASH1L
226501_at		
226544_x_at	biogenesis of lysosomal organelles complex-1, subunit 5, muted	BLOC1S5
226625_at	transforming growth factor, beta receptor III	TGFBR3
226690_at	adenylate cyclase activating polypeptide 1 (pituitary) receptor type I	ADCYAP1R1
226718_at	adhesion molecule with Ig-like domain 1	AMIGO1
226886_at	glutamine–fructose-6-phosphate transaminase 1	GFPT1
226895_at	nuclear factor I/C (CCAAT-binding transcription factor)	NFIC
226898_s_at	splicing factor proline/glutamine-rich	SFPQ
226999_at	RNA-binding region (RNP1, RRM) containing 3	RNPC3
227066_at	MOB kinase activator 3C	MOB3C
227084_at	dystrobrevin, alpha	DTNA
227435_at	KIAA2018	KIAA2018
227527_at	lysine (K)-specific methyltransferase 2D	KMT2D
227556_at	NME/NM23 family member 7	NME7
227663_at	Data not found	
227772_at	Data not found	
227792_at	inositol 1,4,5-trisphosphate receptor interacting protein-like 2	ITPRIPL2
227798_at	SMAD family member 1	SMAD1
227802_at	RUN and FYVE domain containing 3	RUFY3
227847_at	EPM2A (laforin) interacting protein 1	EPM2AIP1
227891_s_at	TAF15 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 68kDa	TAF15
227944_at	protein tyrosine phosphatase, non-receptor type 3	PTPN3
227959_at	Data not found	
228007_at	centrosomal protein 85kDa-like	CEP85L
228287_at	inhibitor of growth family, member 5	ING5
228487_s_at	Data not found	
228556_at	YTH domain containing 1	YTHDC1
228594_at	NAD kinase 2, mitochondrial	NADK2
228662_at	suppressor of cytokine signaling 7	SOCS7
228839_s_at	uncharacterized LOC642361	LOC642361
228855_at	nudix (nucleoside diphosphate linked moiety	
228974_at	Data not found	
229065_at	solute carrier family 35, member F3	SLC35F3
229143_at	CCR4-NOT transcription complex, subunit 3	CNOT3
229145_at	anaphase promoting complex subunit 16	ANAPC16
229264_at	Data not found	
229309_at	adrenoceptor beta 1	ADRB1
229315_at	Data not found	
229353_s_at	nuclear casein kinase and cyclin-dependent kinase substrate 1	NUCKS1
229374_at	EPH receptor A4	EPHA4
229497_at	ankyrin repeat and death domain containing 1A	ANKDD1A
229531_at	Data not found	

The list of probes resulted from Coloured (α, β)-k Feature Set and Generalised (α, β)-k Feature Set problem approach (continued)

Probe ID	Gene	Symbol
229537_at	LIM domain only 4	LMO4
229687_s_at	hypothetical protein LOC100287017	LOC100287017
229725_at	acyl-CoA synthetase long-chain family member 6	ACSL6
229966_at	EWS RNA-binding protein 1	EWSR1
230255_at	gamma-aminobutyric acid (GABA) A receptor, delta	GABRD
230280_at	tripartite motif containing 9	TRIM9
230296_at	chromosome 16 open reading frame 52	C16orf52
230379_x_at	NADH dehydrogenase (ubiquinone) complex I, assembly factor 7	NDUFAF7
230392_at	Data not found	
230621_at	isoamyl acetate-hydrolyzing esterase 1 homolog (S. cerevisiae)	IAH1
230656_s_at	cirrhosis, autosomal recessive 1A (cirhin)	CIRH1A
230713_at	Data not found	
230790_x_at	Data not found	
230885_at	spastic paraplegia 7 (pure and complicated autosomal recessive)	SPG7
230923_at	family with sequence similarity 19 (chemokine (C-C motif)-like), member A1	FAM19A1
231281_at	Data not found	
231329_at	Data not found	
231387_at	hypothetical protein LOC100289494	LOC100289494
231530_s_at	chromosome 11 open reading frame 1	C11orf1
231986_at	regulating synaptic membrane exocytosis 1	RIMS1
232012_at	calpain 1, (mu/I) large subunit	CAPN1
232173_at	C-type lectin domain family 2, member L	CLEC2L
232215_x_at	proline rich 11	PRR11
232288_at	In multiple Geneids	
232311_at	beta-2-microglobulin	B2M
232386_at	vacuolar protein sorting 13 homolog C (S. cerevisiae)	VPS13C
232653_at	Data not found	
232735_at	ankyrin repeat domain 34A	ANKRD34A
233216_at	zinc finger, DHHC-type containing 21	ZDHHC21
233313_at	Data not found	
233319_x_at	Data not found	
233405_at	Data not found	
233437_at	gamma-aminobutyric acid (GABA) A receptor, alpha 4	GABRA4
233647_s_at	cytidine and dCMP deaminase domain containing 1	CDADC1
234163_at	ubiquitin protein ligase E3A	UBE3A
234578_at	Data not found	
234723_x_at	Data not found	
234969_s_at	enhancer of polycomb homolog 1 (Drosophila)	EPC1
234981_x_at	carboxymethylenebutenolidase homolog (Pseudomonas)	CMBL
234989_at	Data not found	
234997_x_at	Data not found	
235049_at	adenylate cyclase 1 (brain)	ADCY1
235200_at	zinc finger protein 561	ZNF561
235288_at	Data not found	
235461_at	tet methylcytosine dioxygenase 2	TET2
235562_at	chromosome 3 open reading frame 70	C3orf70
235567_at	RAR-related orphan receptor A	RORA
235789_at	lysine (K)-specific demethylase 4B	KDM4B
235887_at	Data not found	
235964_x_at	Data not found	
236139_at	Data not found	
236265_at	Sp4 transcription factor	SP4
236268_at	SEC22 vesicle trafficking protein homolog C (S. cerevisiae)	SEC22C

The list of probes resulted from Coloured (α, β)-k Feature Set and Generalised (α, β)-k Feature Set problem approach (continued)

Probe ID	Gene	Symbol
236274_at	eukaryotic translation initiation factor 3, subunit B	EIF3B
236283_x_at	p21 protein (Cdc42/Rac)-activated kinase 2 pseudogene	LOC646214
236338_at	Data not found	
236429_at	zinc finger protein 83	ZNF83
236484_at	Data not found	
236488_s_at	Data not found	
236752_at	Data not found	
236783_at	Kv channel interacting protein 4	KCNIP4
237040_at	CWF19-like 2, cell cycle control (S. pombe)	CWF19L2
237600_at	Data not found	
237768_x_at	Data not found	
237798_at	Data not found	
238017_at	short chain dehydrogenase/reductase family 16C, member 5	SDR16C5
238115_at	DnaJ (Hsp40) homolog, subfamily C, member 18	DNAJC18
238130_at	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 2 interacting protein	NFATC2IP
238147_at	tripartite motif containing 46	TRIM46
238346_s_at	trimethylguanosine synthase 1	TGS1
238470_at	Sys1 golgi trafficking protein	SYS1
238525_at	DEAD (Asp-Glu-Ala-Asp) box helicase 56	DD
238558_at	Data not found	
238560_at	calcium binding and coiled-coil domain 2	CALCOCO2
238602_at	DIS3 like 3'-5' exoribonuclease 2	DIS3L2
238716_at	family with sequence similarity 85, member A	FAM85A
238736_at	REV3-like, polymerase (DNA directed), zeta, catalytic subunit	REV3L
238761_at	ELK4, ETS-domain protein (SRF accessory protein 1)	ELK4
238812_at	Data not found	
238851_at	ankyrin repeat domain 13A	ANKRD13A
238919_at	Data not found	
239026_x_at	ArfGAP with GTPase domain, ankyrin repeat and PH domain 3	AGAP3
239035_at	methylenetetrahydrofolate reductase (NAD(P)H)	MTHFR
239629_at	CASP8 and FADD-like apoptosis regulator	CFLAR
239678_at	Data not found	
239757_at	zinc finger, AN1-type domain 6	ZFAND6
239831_at	transmembrane protein 106C	TMEM106C
240532_at	solute carrier family 32 (GABA vesicular transporter), member 1	SLC32A1
240554_at	A kinase (PRKA) anchor protein 8-like	AKAP8L
240602_at	HBS1-like translational GTPase	HBS1L
241389_at	cholinergic receptor, nicotinic, beta 2 (neuronal)	CHRNB2
241391_at	Data not found	
241727_x_at	dihydrofolate reductase-like 1	DHFRL1
241741_at	cardiolipin synthase 1	CRLS1
241954_at	farnesyl-diphosphate farnesyltransferase 1	FDFT1
242136_x_at	C-terminal binding protein 2 pseudogene	MGC70870
242195_x_at	numb homolog (Drosophila)-like	NUMBL
242319_at	diacylglycerol kinase, gamma 90kDa	DGKG
242343_x_at	Data not found	
242407_at	Data not found	
242413_at	Data not found	
242578_x_at	solute carrier family 22 (organic cation transporter), member 3	SLC22A3
242688_at	Data not found	
242712_x_at	In multiple Geneids	
242829_x_at	F-box and leucine-rich repeat protein 3	FB
242837_at	serine/arginine-rich splicing factor 4	SRSF4

The list of probes resulted from Coloured (α, β)-*k* Feature Set and Generalised (α, β)-*k* Feature Set problem approach (continued)

Probe ID	Gene	Symbol
242973_at	calcium channel, voltage-dependent, L type, alpha 1C subunit	CACNA1C
243318_at	DDB1 and CUL4 associated factor 8	DCAF8
243617_at	zinc finger protein 827	ZNF827
243648_at	zinc finger, BED-type containing 6	ZBED6
244248_at	tetratricopeptide repeat domain 27	TTC27
31874_at	growth arrest-specific 2 like 1	GAS2L1
32069_at	NEDD4 binding protein 1	N4BP1
33148_at	zinc finger RNA binding protein	ZFR
35436_at	golgin A2	GOLGA2
36129_at	small G protein signaling modulator 2	SGSM2
36711_at	v-maf avian musculoaponeurotic fibrosarcoma oncogene homolog F	MAFF
37012_at	capping protein (actin filament) muscle Z-line, beta	CAPZB
37996_s_at	dystrophia myotonica-protein kinase	DMPK
38710_at	OTU deubiquitinase, ubiquitin aldehyde binding 1	OTUB1
40569_at	myeloid zinc finger 1	MZF1
41220_at	septin 9	SEPT9
43427_at	acetyl-CoA carboxylase beta	ACACB
43544_at	mediator complex subunit 16	MED16
46665_at	sema domain, immunoglobulin domain (Ig), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 4C	SEMA4C
49077_at	protein phosphatase methylesterase 1	PPME1
49452_at	acetyl-CoA carboxylase beta	ACACB
51774_s_at	hypothetical protein LOC222070	LOC222070
63825_at	abhydrolase domain containing 2	ABHD2

Probe Id is the list of probes resulted from Coloured (α, β)-*k* Feature Set problem approach. **Gene** is the corresponding gene name. **Symbol** is the corresponding gene symbol.

The list of pathways for the Coloured (α, β)-*k* Feature Set problem approach resulted genes.

Table 8.10: **The list of pathways resulted from Coloured (α, β)-*k* Feature Set problem approach**

Name	EASE score	Genes
Metabolism of Complex Lipids	2.82E-06	DGKG; INPP4A; LPL; SGPL1; TPI1; YWHAZ
Amino Acid Metabolism	2.99E-06	ACAA1; ADSL; GFPT1; GOT2; GSS; PDHA1
Folate biosynthesis	3.48E-06	DHFR
Cell Growth and Death	1.00E-05	ESPL1; PTPN2; PTPN3; PTPRD; RB1
Carbohydrate Metabolism	1.31E-05	ACACB; ALDOA; GRHPR; IDH3A; IDH3B; IDH3G; ME3; PDHA1; PFKFB3; PFKM; PRPS1; TPI1
Valine, leucine and isoleucinebiosynthesis	1.76E-05	PDHA1
Glyoxylate and dicarboxylatemetabolism	1.88E-05	GRHPR
Lipid Metabolism	1.89E-05	ACAA1; ACACB; HSD17B7; NQO2
Alzheimer's disease	8.11E-05	LPL; SNCA
Galactose metabolism	8.11E-05	PFKM

The list of pathways resulted from Coloured (α, β)-k Feature Set problem approach (continued)

Name	EASE score	Genes
Citrate cycle (TCA cycle)	8.11E-05	IDH3A; IDH3B; IDH3G
ATP synthesis	1.88E-04	ATP5G1; ATP5J2; ATP6V1F
Cell Communication	1.88E-04	CAV2; SORBS1
MAPKsignaling pathway	5.75E-04	EGFR
Glutathione metabolism	6.81E-04	GSS
Pentose phosphate pathway	7.16E-04	ALDOA; PFKM; PRPS1
Arginine and proline metabolism	7.42E-04	GOT2
Pyruvate metabolism	1.09E-03	ACACB; GRHPR; ME3; PDHA1
Huntington's disease	1.15E-03	CREBBP
Bile acid biosynthesis	1.21E-03	ACAA1
Sorting and Degradation	1.42E-03	ANAPC5; CDC16; CUL3; PSMB2; PSMB3; PSMB6; PSMC2; PSMD14; TCEB2
Inositol phosphatemetabolism	1.59E-03	INPP4A
Benzoate degradation viahydroxylation	1.63E-03	ACAA1
Signal Transduction	2.20E-03	DGKG; EGFR; INPP4A; ITPR2; PRKCA; PRKCE; PRKCZ; PTPN2; PTPN3; PTPRD
Butanoate metabolism	3.09E-03	PDHA1
Oxidative phosphorylation	3.19E-03	ATP5G1; ATP5J2; ATP6V1F; NDUFA10; NDUFA7; NDUFA9; NDUFS6; UQCRB
Fructose and mannose metabolism	3.32E-03	ALDOA; PFKFB3; PFKM; TPI1
Parkinson's disease	3.36E-03	SNCA
Metabolism of Other Amino Acids	3.41E-03	GSS
Phospholipid degradation	3.54E-03	YWHAZ
Glutamate metabolism	3.61E-03	GFPT1; GOT2; GSS
Purine metabolism	3.78E-03	ADCY1; ADSL; AK3; ENTPD6; GUCY1A3; NME7; PRPS1
Glycerolipid metabolism	3.78E-03	DGKG; LPL; TPI1; YWHAZ
RNA polymerase	5.75E-03	POLR2C
Biosynthesis of Secondary Metabolites	5.75E-03	ACACB; GOT2
Cysteine metabolism	5.75E-03	GOT2
Inositol metabolism	5.75E-03	TPI1
One carbon pool byfolate	5.75E-03	DHFR
Fatty acid biosynthesis (path 2)	5.75E-03	ACAA1
Apoptosis	5.75E-03	PTPN2; PTPN3; PTPRD
Sphingoglycolipidmetabolism	5.75E-03	SGPL1
Phenylalanine metabolism	5.75E-03	GOT2
Riboflavinmetabolism	5.80E-03	PTPN2; PTPN3; PTPRD
Fatty acid metabolism	5.85E-03	ACAA1
Valine, leucine and isoleucinedegradation	5.87E-03	ACAA1
Pyrimidine metabolism	5.92E-03	DUT; ENTPD6; NME7
Basal transcriptionfactors	5.97E-03	GTF2I
Nucleotide Metabolism	6.02E-03	ADCY1; ADSL; AK3; DUT; ENTPD6; GUCY1A3; NME7; PRPS1
Androgen and estrogen metabolism	6.05E-03	HSD17B7
Biodegradation of Xenobiotics	6.11E-03	ACAA1; PTPN2; PTPN3; PTPRD
Prion disease	6.23E-03	LAMC1; NFE2L2
Phosphatidylinositol signaling system	6.32E-03	DGKG; INPP4A; ITPR2; PRKCA; PRKCE; PRKCZ; PTPN2; PTPN3; PTPRD
Carbon fixation	6.58E-03	ALDOA; GOT2; ME3; TPI1
Ubiquitinmediated proteolysis	6.64E-03	ANAPC5; CDC16; CUL3; TCEB2
Fatty acid biosynthesis (path 1)	6.74E-03	ACACB
Aminosugarsmetabolism	6.94E-03	GFPT1

The list of pathways resulted from Coloured (α, β)-*k* Feature Set problem approach (continued)

Name	EASE score	Genes
Propanoate metabolism	7.11E-03	ACACB
Transcription	8.27E-03	GTF2I; POLR2C
Prostaglandin andleukotriene metabolism	8.58E-03	YWHAZ
Alanine and aspartate metabolism	9.50E-03	ADSL; GOT2
Tyrosine metabolism	1.11E-02	GOT2
Tetracyclinebiosynthesis	5.75E-02	ACACB
Phenylalanine, tyrosine andtryptophan biosynthesis	5.75E-02	GOT2
Sterol biosynthesis	5.75E-02	NQO2
Alkaloidbiosynthesis I	5.75E-02	GOT2
Energy Metabolism	5.75E-02	ALDOA; ATP5G1; ATP5J2; ATP6V1F; GOT2; ME3; NDUFA10; NDUFA7; NDUFA9; NDUFS6; TPI1; UQCRB
Metabolism of Cofactors and Vitamins	5.75E-02	BLVRA; BLVRB; DHFR; PTPN2; PTPN3; PTPRD; UROS
Porphyrin andchlorophyll metabolism	1.00E+00	BLVRA; BLVRB; UROS
Glycolysis / Gluconeogenesis	1.00E+00	ALDOA; PDHA1; PFKM; TPI1
Metabolism of Complex Carbohydrates	1.00E+00	GFPT1; MGAT4B
Ribosome	1.00E+00	RPL15; RPL17; RPL3; RPL4; RPL7A
Neurodegenerative Disorders	1.00E+00	CREBBP; LAMC1; LPL; NFE2L2; SNCA
Cell cycle	1.00E+00	ESPL1; PTPN2; PTPN3; PTPRD; RB1
Proteasome	1.00E+00	PSMB2; PSMB3; PSMB6; PSMC2; PSMD14
Translation	1.00E+00	RPL15; RPL17; RPL3; RPL4; RPL7A
N-Glycansbiosynthesis	1.00E+00	MGAT4B
gamma Hexachlorocyclohexane degradation	1.00E+00	PTPN2; PTPN3; PTPRD
integrin mediated celladhesion	1.00E+00	CAV2; SORBS1

Name is the list of resulted pathways. **EASE score** is the corresponding EASE score. **Genes** is the list of genes that are involved in that particular pathway.

The list of pathways for the generalised (α, β)-*k* Feature Set problem approach resulted genes.

Table 8.11: **The list of pathways resulted from Generalised (α, β)-*k* Feature Set problem approach**

Name	EASE score	Genes
Folate biosynthesis	6.38E-08	DHFR
Carbohydrate Metabolism	0.00000059	ABAT; ACACB; AKR1B1; GRHPR; IDH3A; IDH3B; ME3; PDHA1; PFKM; PRPS1; TPI1
Valine, leucine and isoleucinebiosynthesis	0.000000731	PDHA1
Glyoxylate and dicarboxylatemetabolism	0.000000731	GRHPR
MAPKsignaling pathway	0.000000884	EGFR
Lipid Metabolism	0.00000141	ACAA1; ACACB; HSD17B7; NQO2
Alzheimer's disease	0.0000015	GNG3; LPL; SNCA
Galactose metabolism	0.00000176	AKR1B1; PFKM

The list of pathways resulted from Generalised (α, β)-k Feature Set problem approach (continued)

Name	EASE score	Genes
Citrate cycle (TCA cycle)	0.0000021	IDH3A; IDH3B
ATP synthesis	0.0000021	ATP5B; ATP5G1; ATP5J2; ATP6V0B; ATP6V1E1; ATP6V1F
Cell Communication	0.0000021	CAPNS1; SORBS1; TLN2
Glutathione metabolism	0.00000214	GSS
Pentose phosphate pathway	0.00000223	PFKM; PRPS1
Amino Acid Metabolism	0.00000223	ABAT; ACAA1; ADSL; EPRS; GFPT1; GOT2; GSS; PDHA1
Arginine and proline metabolism	0.00000245	EPRS; GOT2
Signal Transduction	0.00000266	DGKG; EGFR; INPP4A; PRKCA; PRKCE; PRKCZ; PTPN3; PTPRD; PTPRM
Butanoate metabolism	0.0000273	ABAT; PDHA1
Oxidative phosphorylation	0.0000287	ATP5B; ATP5G1; ATP5J2; ATP6V0B; ATP6V1E1; ATP6V1F; COX5B; NDUFA10; NDUFA9; NDUFS6; UQCRB; UQCRC2
Cell cycle	0.0000327	ESPL1; PTPN3; PTPRD; PTPRM; RB1
Fructose and mannose metabolism	0.0000338	AKR1B1; PFKM; TPI1
Parkinson's disease	0.0000354	SNCA
Metabolism of Other Amino Acids	0.0000354	ABAT; GSS
Neurodegenerative Disorders	0.0000354	CREBBP; GNG3; LAMC1; LPL; NFE2L2; SNCA
Glutamate metabolism	0.0000354	ABAT; EPRS; GFPT1; GOT2; GSS
Purine metabolism	0.0000354	ADCY1; ADSL; AK3; ENTPD6; GUCY1A3; IMPDH2; NME7; PRPS1
Glycerolipid metabolism	0.0000354	AKR1B1; DGKG; LPL; PAFAH1B1; TPI1; YWHAZ
Pentose and glucuronateinterconversions	0.0000354	AKR1B1
Pyruvate metabolism	0.0000368	ACACB; AKR1B1; GRHPR; ME3; PDHA1
Huntington's disease	0.0000372	CREBBP
Bile acid biosynthesis	0.0000467	ACAA1
Immune System	0.0000467	PLG
Sorting and Degradation	0.0000467	ANAPC5; CDC16; CUL3; PSMB2; PSMB3; PSMC2; PSMD14; TCEB2
Inositol phosphatemetabolism	0.0000467	INPP4A
Benzoate degradation viahydroxylation	0.0000467	ACAA1
Phospholipid degradation	0.000047	YWHAZ
RNA polymerase	0.000047	POLR2C
Riboflavinmetabolism	0.000047	PTPN3; PTPRD; PTPRM
Fatty acid metabolism	0.0000642	ACAA1
Valine, leucine and isoleucinedegradation	0.0000643	ACAA1
Complement andcoagulation cascades	0.0000644	PLG
Pyrimidine metabolism	0.0000679	DTYMK; DUT; ENTPD6; NME7
Basal transcriptionfactors	0.0000679	GTF2I
Nucleotide Metabolism	0.0000679	ADCY1; ADSL; AK3; DTYMK; DUT; ENTPD6; GUCY1A3; IMPDH2; NME7; PRPS1
Androgen and estrogen metabolism	0.0000679	HSD17B7
Biodegradation of Xenobiotics	0.0000679	ACAA1; PTPN3; PTPRD; PTPRM
Prion disease	0.0000698	LAMC1; NFE2L2
Phosphatidylinositol signaling system	0.0000711	DGKG; INPP4A; PRKCA; PRKCE; PRKCZ; PTPN3; PTPRD; PTPRM
Carbon fixation	0.0000724	GOT2; ME3; TPI1

The list of pathways resulted from Generalised (α, β)-*k* Feature Set problem approach (continued)

Name	EASE score	Genes
Ubiquitinmediated proteolysis	0.0000724	ANAPC5; CDC16; CUL3; TCEB2
Fatty acid biosynthesis (path 1)	0.0000724	ACACB
Biosynthesis of Secondary Metabolites	0.0000814	ACACB; GOT2
Apoptosis	0.000085	NFKBIA; PTPN3; PTPRD; PTPRM
Aminosugarsmetabolism	0.000724	GFPT1
Propanoate metabolism	0.000731	ABAT; ACACB
Transcription	0.000736	GTF2I; POLR2C
Metabolism of Complex Lipids	0.000738	AKR1B1; DGKG; INPP4A; LPL; PAFAH1B1; SGPL1; TPI1; YWHAZ
Prostaglandin andleukotriene metabolism	0.000739	YWHAZ
Alanine and aspartate metabolism	0.000766	ABAT; ADSL; GOT2
Tyrosine metabolism	0.000786	GOT2
Glycolysis / Gluconeogenesis	0.000992	PDHA1; PFKM; TPI1
Cysteine metabolism	0.00821	GOT2
Translation	0.00821	EPRS; RPL15; RPL3
Inositol metabolism	0.00821	TPI1
One carbon pool byfolate	0.00821	DHFR
Ribosome	0.00821	RPL15; RPL3
Fatty acid biosynthesis (path 2)	0.00821	ACAA1
Cell Growth and Death	0.00831	ESPL1; NFKBIA; PTPN3; PTPRD; PTPRM; RB1
Sphingoglycolipidmetabolism	0.0085	SGPL1
Phenylalanine metabolism	0.00854	GOT2
Tetracyclinebiosynthesis	0.00864	ACACB
Phenylalanine, tyrosine andtryptophan biosynthesis	0.088	GOT2
Proteasome	0.0884	PSMB2; PSMB3; PSMC2; PSMD14
Sterol biosynthesis	0.0925	NQO2
Alkaloidbiosynthesis I	0.0925	GOT2
Energy Metabolism	0.0925	ATP5B; ATP5G1; ATP5J2; ATP6V0B; ATP6V1E1; ATP6V1F; COX5B; GOT2; ME3; NDUFA10; NDUFA9; NDUFS6; TPI1; UQCRB; UQCRC2
Metabolism of Complex Carbohydrates	0.0958	GFPT1; MGAT4B
Metabolism of Cofactors and Vitamins	0.0959	BLVRA; BLVRB; DHFR; EPRS; PTPN3; PTPRD; PTPRM; UROS
beta Alanine metabolism	0.0974	ABAT
N Glycansbiosynthesis	0.0974	MGAT4B
gamma Hexachlorocyclohexane degradation	0.0974	PTPN3; PTPRD; PTPRM
Porphyrin andchlorophyll metabolism	0.0992	BLVRA; BLVRB; EPRS; UROS
Integrin mediated celladhesion	0.974	CAPNS1; SORBS1; TLN2
aminoacyl tRNAbiosynthesis	0.974	EPRS

Name is the list of resulted pathways. **EASE score** is the corresponding EASE score. **Genes** is the list of genes that are involved in that particular pathway.

Comparison result of RankProd and Coloured (α, β)-*k* Feature Set problem approach.

Table 8.12: **Comparison result of RankProd with Coloured (α, β)-*k* Feature Set and Generalised (α, β)-*k* Feature Set problem approach.**

Probe ID	RP	RP/RSum
1558678_s_at	Up	14.0555
229353_s_at	Up	32.8452
227084_at	Up	108.7779
211921_x_at	Up	135.3077
203485_at	Up	148.4955
221476_s_at	Down	161.8223
232215_x_at	Up	209.3067
208835_s_at	Up	222.8763
209251_x_at	Down	245.2321
234981_x_at	Up	246.5703
225239_at	Up	258.0749
212581_x_at	Down	274.1996
203752_s_at	Up	276.5254
227556_at	Down	320.4099
234723_x_at	Up	356.1144
208942_s_at	Up	361.3957
213453_x_at	Down	367.3052
1553186_x_at	Up	383.3695
205751_at	Down	387.1836
217398_x_at	Down	401.8526
204720_s_at	Down	429.3097
214394_x_at	Up	429.9386
225649_s_at	Up	437.78
200639_s_at	Down	442.0173
209225_x_at	Up	454.2189
230296_at	Up	462.7402
224568_x_at	Up	467.7329
222982_x_at	Up	482.2697
226086_at	Down	527.6699
201305_x_at	Up	568.5796
1568612_at	Down	588.7767
200640_at	Down	594.6262
221506_s_at	Up	610.0088
41220_at	Up	630.1333
201217_x_at	Down	631.5344
201991_s_at	Up	657.3275
31874_at	Up	665.8403
203909_at	Down	670.0062
222380_s_at	Up	674.09
209169_at	Up	677.4073
234989_at	Up	677.4917
206140_at	Up	710.0032
215600_x_at	Up	768.4976
209250_at	Up	780.0573
242195_x_at	Up	786.6162
202712_s_at	Down	795.6037
226895_at	Up	799.5595
203140_at	Up	800.0041
222043_at	Up	814.1865
216524_x_at	Up	830.2956
233702_x_at	Up	838.3096
213545_x_at	Down	859.4138
204229_at	Down	870.4442

Comparison result of RankProd with Coloured (α, β)-*k* Feature Set and Generalised (α, β)-*k* Feature Set problem approach. (continued)

Probe ID	RP	RP/RSum
228662_at	Up	876.9168
201065_s_at	Up	878.5218
46665_at	Up	887.3926
210825_s_at	Down	902.5565
208549_x_at	Up	918.4285
224628_at	Down	937.396
49452_at	Up	947.2932
242239_at	Up	964.8067
210976_s_at	Down	989.8563
200708_at	Down	1019.5751
214594_x_at	Up	1020.941
226447_at	Up	1051.1477
219028_at	Up	1053.8601
36129_at	Up	1055.5393
201410_at	Down	1075.6059
220609_at	Up	1080.6593
202178_at	Down	1083.8877
234969_s_at	Up	1084.6926
36711_at	Up	1089.7614
220071_x_at	Up	1111.3633
200822_x_at	Down	1121.1827
202160_at	Up	1143.1558
201441_at	Down	1156.9447
212265_at	Up	1186.6332
217927_at	Down	1225.856
242829_x_at	Up	1227.9716
214375_at	Up	1241.2239
201828_x_at	Down	1243.5453
204466_s_at	Down	1286.3477
223134_at	Up	1290.0063
221488_s_at	Down	1304.1046
205257_s_at	Down	1305.5117
215963_x_at	Down	1320.4211
212177_at	Up	1347.3329
212372_at	Down	1353.2277
202961_s_at	Down	1357.8652
225923_at	Down	1372.5845
1561346_at	Up	1424.1319
203509_at	Down	1434.7709
212296_at	Down	1438.994
210111_s_at	Up	1441.4453
200976_s_at	Down	1442.0211
209289_at	Up	1454.9565
205353_s_at	Down	1466.6575
228594_at	Up	1516.3553
236783_at	Down	1539.0722
225240_s_at	Up	1551.0981
214743_at	Up	1578.3389
216550_x_at	Up	1591.722
212114_at	Down	1592.8844
227847_at	Down	1635.5017
215091_s_at	Down	1647.8373
209001_s_at	Down	1668.6804
200754_x_at	Down	1682.2226

Comparison result of RankProd with Coloured (α, β)-*k* Feature Set and Generalised (α, β)-*k* Feature Set problem approach. (continued)

Probe ID	RP	RP/RSum
208859_s_at	Up	1706.4235
230923_at	Down	1720.4901
43427_at	Up	1723.3154
227527_at	Up	1735.4662
225571_at	Up	1756.4402
208704_x_at	Down	1781.8483
211993_at	Up	1782.8699
243791_at	Up	1793.423
220966_x_at	Down	1810.5661
1316_at	Up	1828.4984
203431_s_at	Down	1840.9345
210679_x_at	Up	1858.6604
212451_at	Up	1859.8691
207789_s_at	Down	1886.5229
217871_s_at	Down	1946.5334
210686_x_at	Up	1959.7108
214707_x_at	Up	2039.3322
200098_s_at	Down	2040.5294
213089_at	Up	2049.8047
218330_s_at	Up	2063.0806
242578_x_at	Up	2067.1215
207081_s_at	Down	2075.082
213808_at	Down	2113.2949
201527_at	Down	2136.0197
225092_at	Down	2151.7633
209343_at	Up	2168.887
242343_x_at	Up	2179.9858
235964_x_at	Up	2198.8075
239179_at	Up	2202.5911
1555495_a_at	Up	2214.1863
203606_at	Down	2223.421
215383_x_at	Up	2240.616
219387_at	Up	2243.0554
212501_at	Up	2244.6017
214246_x_at	Up	2266.5547
225501_at	Up	2304.9933
222154_s_at	Down	2322.6066
220942_x_at	Down	2323.0839
212450_at	Up	2339.9954
209583_s_at	Down	2351.2612
222651_s_at	Up	2361.9197
209258_s_at	Up	2362.4058
200625_s_at	Down	2370.1612
208732_at	Down	2392.8281
206562_s_at	Up	2421.1817
207966_s_at	Up	2435.092
201371_s_at	Down	2436.1717
229497_at	Up	2470.0995
200693_at	Down	2472.8104
223673_at	Up	2477.9117
208972_s_at	Down	2487.0845
203297_s_at	Up	2487.4001
49077_at	Down	2503.9184
200039_s_at	Down	2532.9213

Comparison result of RankProd with Coloured (α, β)-*k* Feature Set and Generalised (α, β)-*k* Feature Set problem approach. (continued)

Probe ID	RP	RP/RSum
224366_s_at	Down	2534.5756
214799_at	Up	2537.6596
222047_s_at	Up	2541.5133
200085_s_at	Down	2551.978
204584_at	Up	2552.4649
223035_s_at	Down	2553.5102
212468_at	Up	2566.2028
244610_x_at	Up	2587.2832
217761_at	Up	2635.229
220081_x_at	Up	2638.9223
214043_at	Down	2687.7035
209558_s_at	Up	2694.744
201400_at	Down	2695.0841
233319_x_at	Up	2708.0047
226544_x_at	Up	2717.9965
201704_at	Down	2756.4248
203956_at	Up	2773.3075
233313_at	Up	2808.2115
226143_at	Up	2808.5204
241786_at	Up	2809.6744
212483_at	Up	2812.8879
208710_s_at	Up	2813.4318
205559_s_at	Up	2815.9153
218026_at	Down	2819.6046
221853_s_at	Down	2869.0207
242240_at	Up	2872.9226
230885_at	Up	2893.7268
228007_at	Up	2934.7385
219032_x_at	Down	2948.3193
217713_x_at	Up	2953.3995
223319_at	Down	2959.3823
204786_s_at	Up	2990.4158
236752_at	Up	2991.3574
234163_at	Up	3001.5349
236488_s_at	Up	3001.5888
212462_at	Up	3027.8204
242688_at	Up	3037.1577
201290_at	Up	3061.3041
225516_at	Up	3069.028
238558_at	Up	3087.7043
229309_at	Up	3092.4821
217939_s_at	Down	3101.9653
237768_x_at	Up	3128.9727
213015_at	Up	3164.6525
226625_at	Up	3220.9923
209023_s_at	Up	3257.6288
225356_at	Up	3288.3434
203146_s_at	Down	3329.8739
212104_s_at	Down	3339.6844
202070_s_at	Down	3378.2475
227802_at	Up	3411.0758
226999_at	Up	3426.8866
231281_at	Up	3428.2976
202829_s_at	Down	3439.2194

Comparison result of RankProd with Coloured (α, β)-k Feature Set and Generalised (α, β)-k Feature Set problem approach. (continued)

Probe ID	RP	RP/RSum
213457_at	Up	3450.47
217860_at	Down	3477.2505
225636_at	Up	3480.0972
203068_at	Up	3501.4537
223380_s_at	Up	3510.0167
1558695_at	Up	3514.0159
201500_s_at	Down	3528.8735
220316_at	Up	3530.7003
222024_s_at	Up	3551.7801
226169_at	Up	3558.5976
226112_at	Up	3568.4585
239629_at	Up	3576.0368
203773_x_at	Down	3577.1319
211930_at	Up	3581.3127
225917_at	Up	3592.223
208969_at	Down	3607.586
223480_s_at	Down	3615.2681
203132_at	Up	3629.0475
205787_x_at	Up	3636.9375
229145_at	Up	3656.425
201145_at	Down	3668.2232
224774_s_at	Up	3669.6183
230255_at	Down	3682.5013
209609_s_at	Down	3715.5421
239035_at	Up	3717.1839
218381_s_at	Up	3725.7323
225234_at	Up	3733.297
212832_s_at	Down	3742.6934
200047_s_at	Up	3758.8735
236338_at	Up	3763.3935
225493_at	Up	3771.7048
37012_at	Up	3791.1743
208936_x_at	Down	3817.6154
222513_s_at	Up	3837.1375
202858_at	Down	3864.196
212207_at	Up	3878.7755
204957_at	Down	3882.6057
241893_at	Up	3888.2624
212491_s_at	Down	3889.0422
228287_at	Up	3892.6101
225223_at	Up	3902.5975
227798_at	Up	3902.6559
200980_s_at	Down	3910.3871
213710_s_at	Down	3912.3385
224598_at	Up	3943.6731
213039_at	Up	3969.1318
200685_at	Up	3974.6726
202974_at	Down	3989.1306
214823_at	Down	4014.827
207132_x_at	Down	4022.0544
212252_at	Down	4036.491
208611_s_at	Down	4055.1583
202927_at	Down	4082.6614
224227_s_at	Up	4124.3358

Comparison result of RankProd with Coloured (α, β)-*k* Feature Set and Generalised (α, β)-*k* Feature Set problem approach. (continued)

Probe ID	RP	RP/RSum
237040_at	Up	4125.0582
227792_at	Up	4136.6214
233437_at	Down	4182.7313
223037_at	Down	4197.2167
215021_s_at	Down	4198.516
231986_at	Down	4208.9383
1569302_at	Up	4209.9194
227772_at	Up	4222.8981
202144_s_at	Down	4231.4108
211386_at	Up	4236.2091
217028_at	Up	4236.4535
216101_at	Up	4252.1041
51774_s_at	Up	4253.4869
201415_at	Down	4259.7248
236265_at	Up	4265.396
213744_at	Down	4279.5789
209268_at	Down	4299.8132
224689_at	Down	4305.52
202360_at	Up	4329.2292
217969_at	Down	4362.7917
1559391_s_at	Up	4399.3377
212209_at	Up	4409.2136
218953_s_at	Down	4449.6799
223024_at	Down	4457.6668
210908_s_at	Down	4462.5031
204270_at	Up	4467.9906
219389_at	Down	4471.0909
242712_x_at	Up	4474.9416
1568763_s_at	Up	4477.6442
238736_at	Up	4482.6071
226101_at	Down	4511.349
81811_at	Up	4521.28
209715_at	Up	4537.6003
225750_at	Down	4593.3842
217077_s_at	Down	4603.6119
215385_at	Up	4639.6388
239026_x_at	Up	4665.4506
201570_at	Down	4695.5607
220072_at	Up	4713.159
243561_at	Up	4713.3443
224878_at	Down	4733.734
244778_x_at	Up	4734.8523
226690_at	Up	4788.6507
211752_s_at	Down	4799.6614
203122_at	Down	4811.7757
215543_s_at	Down	4811.9489
204663_at	Down	4819.2987
240758_at	Up	4821.2303
222395_s_at	Up	4838.1538
233816_at	Up	4839.6569
212992_at	Down	4853.557
238851_at	Up	4854.6941
218247_s_at	Up	4866.3851
32069_at	Up	4870.0985

Comparison result of RankProd with Coloured (α, β)-k Feature Set and Generalised (α, β)-k Feature Set problem approach. (continued)

Probe ID	RP	RP/RSum
218048_at	Down	4896.7554
224999_at	Up	4918.7124
216745_x_at	Up	4926.7959
214436_at	Down	4933.3369
224666_at	Down	4982.8911
239831_at	Up	4986.386
1570210_x_at	Up	4993.879
209991_x_at	Down	5002.3528
202120_x_at	Down	5007.5091
227663_at	Up	5048.759
201509_at	Down	5069.9517
212995_x_at	Down	5073.7102
211616_s_at	Down	5086.4741
224739_at	Up	5107.7277
230280_at	Down	5115.6987
228839_s_at	Up	5127.7242
204977_at	Down	5138.3318
201057_s_at	Down	5144.6956
221486_at	Down	5153.3887
200732_s_at	Down	5175.5822
243318_at	Up	5194.9736
221942_s_at	Down	5196.0318
207598_x_at	Up	5213.9562
241391_at	Up	5234.528
209409_at	Down	5244.8715
205967_at	Up	5255.6098
225050_at	Down	5271.91
226037_s_at	Down	5303.8406
219670_at	Down	5319.5352
35436_at	Up	5328.6011
229537_at	Down	5339.4792
215504_x_at	Up	5341.5608
200072_s_at	Down	5357.436
202868_s_at	Down	5384.9795
213517_at	Up	5388.7552
235575_at	Up	5439.1873
202534_x_at	Up	5460.7992
203849_s_at	Up	5469.1894
219019_at	Up	5474.352
242319_at	Up	5476.5637
227435_at	Up	5480.9514
214499_s_at	Up	5493.8844
201067_at	Down	5511.9409
1561657_at	Up	5513.6104
1569200_at	Up	5516.7588
204731_at	Up	5542.183
43544_at	Up	5564.9432
236949_at	Up	5571.0485
40569_at	Up	5577.4924
209440_at	Down	5586.4151
230656_s_at	Down	5590.1901
201434_at	Down	5601.8198
213195_at	Down	5607.5292
225003_at	Down	5619.566

Comparison result of RankProd with Coloured (α, β)-*k* Feature Set and Generalised (α, β)-*k* Feature Set problem approach. (continued)

Probe ID	RP	RP/RSum
200042_at	Down	5647.6253
205117_at	Up	5653.5985
235049_at	Down	5654.2645
217819_at	Down	5654.3446
238919_at	Up	5662.6052
235697_at	Up	5670.5822
217250_s_at	Up	5698.6953
232653_at	Up	5701.7396
212095_s_at	Up	5709.3321
225870_s_at	Down	5790.0714
236268_at	Up	5797.3474
215553_x_at	Up	5801.9793
1560116_a_at	Up	5832.8945
238346_s_at	Up	5843.472
63825_at	Up	5848.5593
221952_x_at	Down	5897.923
219671_at	Down	5903.7264
243329_at	Up	5927.73
243648_at	Up	5928.8164
226718_at	Down	5929.1077
213079_at	Down	5945.4455
209066_x_at	Down	5955.9567
202121_s_at	Down	5956.9898
209553_at	Down	5961.2807
201324_at	Up	5983.887
212228_s_at	Down	6069.5894
217726_at	Down	6083.5486
244859_at	Up	6160.5795
202594_at	Down	6170.9208
1568877_a_at	Up	6177.1634
222364_at	Up	6178.7316
217482_at	Up	6183.8109
1558831_x_at	Up	6214.6546
1557155_a_at	Up	6217.7069
216527_at	Up	6257.9245
214934_at	Up	6302.963
218520_at	Down	6304.3504
201709_s_at	Down	6314.712
212322_at	Up	6327.9051
219961_s_at	Down	6328.7643
241727_x_at	Up	6331.9083
227959_at	Up	6345.2087
225496_s_at	Down	6349.3697
212384_at	Up	6374.2802
224972_at	Down	6399.2817
238560_at	Up	6401.2495
200771_at	Up	6412.6308
230392_at	Up	6443.7773
230713_at	Up	6515.4469
238115_at	Down	6527.5315
207084_at	Up	6544.6669
212178_s_at	Up	6555.7817
201146_at	Up	6567.4821
203814_s_at	Down	6599.3751

Comparison result of RankProd with Coloured (α, β)-*k* Feature Set and Generalised (α, β)-*k* Feature Set problem approach. (continued)

Probe ID	RP	RP/RSum
212432_at	Down	6627.8647
222809_x_at	Up	6661.0355
229531_at	Up	6666.0734
230621_at	Up	6681.5154
202551_s_at	Up	6697.0614
200954_at	Down	6728.7246
244803_at	Up	6734.8637
218271_s_at	Down	6743.6325
202330_s_at	Up	6746.9666
205514_at	Down	6753.5256
201836_s_at	Down	6762.4768
233647_s_at	Down	6769.7338
211779_x_at	Down	6783.0913
220615_s_at	Down	6816.1255
229725_at	Down	6818.7093
241954_at	Up	6819.0407
202135_s_at	Down	6827.5283
239096_at	Up	6838.3426
203113_s_at	Up	6850.3456
235567_at	Up	6881.8506
213421_x_at	Down	6888.0341
229264_at	Up	6910.3235
243827_at	Up	6921.3528
237018_at	Up	6922.9251
215167_at	Down	6925.1651
242407_at	Up	7030.7701
212645_x_at	Down	7043.309
240532_at	Down	7044.6109
222428_s_at	Up	7050.0229
214850_at	Up	7063.2231
212348_s_at	Down	7073.0489
1568603_at	Down	7081.3799
1563881_at	Up	7113.3411
225219_at	Up	7127.6581
218843_at	Down	7139.6652
236139_at	Up	7143.3793
236283_x_at	Up	7149.5909
230790_x_at	Up	7155.1267
206652_at	Up	7161.2566
224862_at	Up	7167.7124
208996_s_at	Down	7185.7976
224151_s_at	Up	7211.9303
236934_at	Up	7243.5054
203031_s_at	Down	7256.5952
203911_at	Up	7324.5021
225055_at	Up	7326.0944
229143_at	Up	7329.6814
209163_at	Down	7331.9223
212088_at	Down	7424.1792
221886_at	Up	7432.7711
1565620_at	Up	7449.1277
204718_at	Down	7464.2942
236901_at	Up	7542.3926
240602_at	Up	7577.867

Comparison result of RankProd with Coloured (α, β)-*k* Feature Set and Generalised (α, β)-*k* Feature Set problem approach. (continued)

Probe ID	RP	RP/RSum
1553693_s_at	Up	7649.6388
229374_at	Down	7650.3736
230379_x_at	Up	7703.7201
239102_s_at	Up	7708.241
229065_at	Down	7737.9546
229687_s_at	Up	7741.1983
229966_at	Up	7754.2649
239497_at	Up	7773.3218
232173_at	Down	7852.1918
221772_s_at	Down	7852.2482
203549_s_at	Down	7867.4418
209177_at	Down	7871.5055
238147_at	Up	7882.9982
225298_at	Down	7893.0947
202201_at	Down	7896.9452
228974_at	Up	7909.7174
201810_s_at	Down	7959.2504
1569482_at	Up	7961.6245
236484_at	Up	8001.3068
202505_at	Down	8007.1172
201636_at	Up	8011.6858
225864_at	Up	8039.5154
209079_x_at	Up	8071.1355
236274_at	Up	8078.2411
209534_x_at	Up	8081.4038
1560741_at	Up	8100.2531
1566480_x_at	Up	8121.0692
1554593_s_at	Down	8121.7013
201622_at	Down	8124.1578
204559_s_at	Down	8159.3066
235200_at	Up	8163.0436
215280_s_at	Up	8186.4665
235288_at	Up	8194.5579
204622_x_at	Up	8250.5244
240554_at	Up	8251.0829
233037_at	Up	8262.71
241216_at	Up	8268.7477
238017_at	Down	8305.923
238761_at	Up	8341.8883
232012_at	Up	8421.6982
244511_at	Up	8439.1825
211964_at	Up	8442.0029
241741_at	Up	8454.9748
228487_s_at	Up	8526.8384
225172_at	Up	8545.5648
201083_s_at	Up	8588.0472
240005_at	Down	8606.4209
201586_s_at	Down	8616.9886
227944_at	Down	8622.3621
220269_at	Down	8634.7429
201182_s_at	Up	8665.3125
231329_at	Up	8700.2699
239333_x_at	Up	8726.4012
238602_at	Up	8729.4825

Comparison result of RankProd with Coloured (α, β)-*k* Feature Set and Generalised (α, β)-*k* Feature Set problem approach. (continued)

Probe ID	RP	RP/RSum
216308_x_at	Down	8797.064
226501_at	Up	8837.2844
242973_at	Up	8842.852
213535_s_at	Down	8889.1354
204517_at	Up	8893.5694
221877_at	Down	8916.6129
243593_s_at	Up	8931.7418
233216_at	Up	8957.7708
212652_s_at	Down	8957.8778
202138_x_at	Down	8958.8654
209686_at	Up	9032.5545
213093_at	Up	9098.7162
229315_at	Up	9110.7275
200053_at	Down	9126.2423
223679_at	Up	9153.2767
225463_x_at	Down	9211.109
209164_s_at	Up	9222.3376
222160_at	Up	9224.8233
1559060_a_at	Up	9232.7332
211464_x_at	Up	9237.9396
200775_s_at	Down	9272.8938
240948_at	Up	9284.3534
238470_at	Up	9335.6846
201548_s_at	Up	9345.6429
203466_at	Down	9363.0934
222745_s_at	Up	9372.0549
205594_at	Up	9374.4055
222284_at	Up	9464.5999
232288_at	Up	9501.5407
234578_at	Up	9532.3606
204552_at	Down	9557.3646
241798_at	Up	9587.7695
239678_at	Down	9664.2445
215493_x_at	Down	9672.5596
226195_at	Down	9688.4593
217937_s_at	Up	9732.0681
225484_at	Down	9765.9282
223185_s_at	Up	9778.1356
212673_at	Down	9799.3543
235730_at	Up	9800.5416
226886_at	Down	9871.0812
1552507_at	Up	9887.9547
232386_at	Up	9943.1059
228556_at	Up	9965.4583
209586_s_at	Down	9989.7307
238716_at	Up	10000.9545
222457_s_at	Up	10043.3782
212896_at	Down	10056.9553
226223_at	Up	10057.908
227891_s_at	Up	10101.7221
221864_at	Up	10126.695
232735_at	Down	10136.7616
217946_s_at	Down	10150.9925
218302_at	Down	10152.3408

Comparison result of RankProd with Coloured (α, β)-*k* Feature Set and Generalised (α, β)-*k* Feature Set problem approach. (continued)

Probe ID	RP	RP/RSum
203082_at	Down	10163.434
204266_s_at	Down	10189.3339
243025_at	Up	10234.8759
1561686_at	Up	10243.413
237600_at	Up	10269.5116
213693_s_at	Up	10273.0413
206547_s_at	Down	10281.6963
1561158_at	Up	10287.9578
234997_x_at	Up	10391.4602
231387_at	Up	10538.9013
244045_at	Up	10624.1216
244027_at	Up	10646.6935
209751_s_at	Down	10669.4654
1556690_s_at	Up	10681.0372
202717_s_at	Down	10718.1799
224369_s_at	Down	10744.2331
210840_s_at	Up	10760.5812
219709_x_at	Down	10766.5931
1558792_x_at	Up	10788.0899
225961_at	Up	10890.5243
222439_s_at	Down	10912.1028
240331_at	Up	10969.9897
224737_x_at	Up	11078.9557
204481_at	Up	11169.092
217904_s_at	Down	11208.925
218543_s_at	Up	11326.308
242413_at	Up	11487.3816
212506_at	Up	11641.3016
38710_at	Up	11758.2878
204528_s_at	Up	11790.3657
223519_at	Up	12126.4061
244248_at	Up	12162.0474
227066_at	Up	12185.1376
238130_at	Up	12231.5759
232311_at	Up	12291.4374
1557293_at	Up	12317.0381
222610_s_at	Up	12336.5085
242837_at	Up	12357.8149
1562280_at	Up	12473.1053
224414_s_at	Up	12485.3424
215698_at	Up	12506.7726
235789_at	Up	12718.0766
235562_at	Up	12740.4199
1562416_at	Up	12748.2276
225655_at	Up	12806.3792
33148_at	Up	12826.6381

Probe ID is the list of probes that are present in the top listed probes in the RankProd. **RP** indicates that the gene is up or down regulated in RankProd result. **RP/RSum** indicates the rank product resulted from RankProd approach.

Comparison result of GeneMeta with Coloured (α, β)-k Feature Set and Generalised (α, β)-k Feature Set problem approach.

Table 8.13: **Comparison result of GeneMeta with Coloured (α, β)-k Feature Set and Generalised (α, β)-k Feature Set problem approach.**

Probe ID	Z-Score	FDR
205967_at	0.002980233	0.600418492
203849_s_at	0.085746594	0.685143169
203911_at	0.120488963	0.667634263
225961_at	0.145670441	0.655080524
212506_at	0.476402242	0.741177558
206122_at	0.586984543	0.706115628
204528_s_at	0.625093052	0.774438135
218298_s_at	1.117676013	0.975011066
222439_s_at	1.22235687	0.929571643
203113_s_at	1.360780964	0.42958515
222428_s_at	1.364718345	0.834915635
209686_at	1.373291549	0.417551355
218865_at	1.396539	0.903287445
200775_s_at	1.519101464	1.012475052
208955_at	1.639640553	0.880625493
213093_at	1.701402075	1.008148768
204680_s_at	1.793965891	0.911502107
38710_at	1.969284212	0.952096374
202025_x_at	2.224504511	1.010088036
220295_x_at	2.304271581	0.151832667
229725_at	2.362476844	1.010668119
202820_at	2.472919076	0.213992628
220136_s_at	2.577461095	1.082456251
217904_s_at	2.661704709	1.06977057
201146_at	2.67861347	0.20687138
205117_at	2.712000252	0.187200188
218522_s_at	2.774960108	1.093226094
203549_s_at	2.816953218	1.065835059
204266_s_at	2.855419739	1.084622207
214934_at	2.906121925	0.289622086
213039_at	2.923567843	0.24320932
211964_at	3.051536981	0.19303222
217250_s_at	3.206594269	0.197610194
209586_s_at	3.357507239	1.059782283
204321_at	3.385726537	0.147291011
226690_at	3.450608344	0.169022392
235461_at	3.457082398	0.118036188
244511_at	3.468882222	0.230406702
200047_s_at	3.473211245	0.204347575
215514_at	3.53362367	0.156028756
243617_at	3.540991357	0.122763297
242973_at	3.572264344	0.136527059
217946_s_at	3.576195593	1.087720183
209164_s_at	3.633170685	0.139053794
209066_x_at	3.654888004	1.039558058
215306_at	3.677882314	0.01107766

Comparison result of GeneMeta with Coloured (α, β)-*k* Feature Set and Generalised (α, β)-*k* Feature Set problem approach. (continued)

Probe ID	Z-Score	FDR
225496_s_at	3.71010996	1.090847536
206273_at	3.737161182	0.091537863
232735_at	3.744423748	1.094815128
201548_s_at	3.769456221	0.068104105
203082_at	3.787232277	1.093858193
231329_at	3.825337656	0.190834251
242136_x_at	3.826321187	0.004998679
201810_s_at	3.893479467	1.089069082
238919_at	3.950537428	0.088773431
215280_s_at	3.964366884	0.088829685
1558279_a_at	3.99285757	0.006586568
201636_at	4.011122039	0.162072223
241389_at	4.021788988	1.080821717
209409_at	4.045057726	1.088831625
214043_at	4.065634924	1.078746484
208710_s_at	4.065826642	0.209810928
212104_s_at	4.069231526	1.087727977
209316_s_at	4.089980451	1.094292087
37996_s_at	4.090562747	0.105283958
230621_at	4.098156844	0.009598939
225484_at	4.118727362	1.073140045
239678_at	4.143599372	1.094762523
215493_x_at	4.146083005	1.078003595
221942_s_at	4.15618037	1.054511762
204584_at	4.188679271	0.185343148
235562_at	4.254622925	0.045097353
244859_at	4.305453609	0.004859043
221486_at	4.307230431	1.092541101
233908_x_at	4.328660366	3.26264E-05
209079_x_at	4.333718001	0.055016134
224369_s_at	4.380419754	1.079489216
207966_s_at	4.385158101	0.07107626
33148_at	4.39486647	0.030165309
211616_s_at	4.427748232	1.093747154
225298_at	4.42875948	1.08769456
224737_x_at	4.440572143	0.025916353
219204_s_at	4.443937075	1.066210555
40569_at	4.461949427	0.089545556
1569200_at	4.469951186	0.010899188
219387_at	4.477445002	0.013447719
200053_at	4.484637591	1.092831027
213298_at	4.496774115	0.066569397
235887_at	4.498454295	1.90658E-05
202594_at	4.514842392	1.090352323
219709_x_at	4.560066483	1.06540911
212673_at	4.589312001	1.094325907
217819_at	4.594795182	1.083278478
212896_at	4.600748389	1.094438552
215889_at	4.603711772	1.47493E-05
204622_x_at	4.615914758	0.005523428
35436_at	4.617025757	0.044373592
210111_s_at	4.630550999	0.157085777
204718_at	4.631617236	1.048111553
232012_at	4.647022851	0.018168087

Comparison result of GeneMeta with Coloured (α, β)-*k* Feature Set and Generalised (α, β)-*k* Feature Set problem approach. (continued)

Probe ID	Z-Score	FDR
206307_s_at	4.651430645	5.48246E-06
201083_s_at	4.657172503	0.016508088
201586_s_at	4.698210538	1.073544099
1560689_s_at	4.730191622	7.95022E-05
243318_at	4.735588028	0.002271039
218543_s_at	4.738249007	0.001938363
234578_at	4.749288449	0.000930889
237798_at	4.828440342	0.014901903
213535_s_at	4.838026001	1.0933447
209534_x_at	4.839668857	0.060779635
235789_at	4.856782379	0.003118053
233647_s_at	4.857749973	1.094948639
238115_at	4.874365025	1.094794781
234081_at	4.916989203	0
238147_at	4.920303058	0.008888889
37012_at	4.92301942	0.057797766
226223_at	4.937684248	0.009157989
212652_s_at	4.943739154	1.07332513
215504_x_at	5.01248784	0.000184385
228855_at	5.014936002	1.019431793
205570_at	5.049161716	0.002291935
43544_at	5.051042995	0.080099907
201324_at	5.05397533	0
240532_at	5.072810184	1.034131983
221864_at	5.075739136	0
221877_at	5.089340904	1.079142557
210825_s_at	5.092611623	1.0803368
204481_at	5.104272303	0
239469_at	5.106947459	0.000742072
226898_s_at	5.151361424	0.002384837
233405_at	5.168060654	0
1563781_at	5.175954033	0
32069_at	5.177447998	0.027279842
238130_at	5.203183562	0
215963_x_at	5.206263332	1.080697707
1557293_at	5.226143251	5.71211E-05
242837_at	5.242002556	5.0702E-05
203485_at	5.242374968	0.000461754
210701_at	5.253612587	0.000245298
1557895_at	5.257719696	0.00035092
200042_at	5.259022209	1.094201362
201217_x_at	5.279934963	1.076036093
202717_s_at	5.293980199	1.079456813
229687_s_at	5.302899013	0.017896892
235288_at	5.307813788	0
225516_at	5.31303473	0
229264_at	5.313912642	0.068392638
238761_at	5.320047597	0.022592819
214394_x_at	5.325113938	0.00088103
225463_x_at	5.327527125	1.056138579
243275_at	5.346668035	0
202505_at	5.357169345	1.082665724
200072_s_at	5.362308418	1.066910954
244673_at	5.367238715	0

Comparison result of GeneMeta with Coloured (α, β)-*k* Feature Set and Generalised (α, β)-*k* Feature Set problem approach. (continued)

Probe ID	Z-Score	FDR
212322_at	5.370760123	0.015406638
220609_at	5.372348469	0.008964624
1562898_at	5.393507944	0
237600_at	5.397367351	0.000877839
242688_at	5.39823512	1.73762E-05
228974_at	5.416529534	0.00415917
232797_at	5.429541864	0.0040095
202138_x_at	5.432093646	1.014715446
200685_at	5.434244989	0.001123682
209440_at	5.437012145	1.077823876
222024_s_at	5.438479557	0.042698503
226886_at	5.452750628	1.082392844
232386_at	5.458064031	0.001189619
36129_at	5.46131904	0.038083629
230379_x_at	5.470450159	9.65251E-05
242407_at	5.470519293	5.05443E-05
222284_at	5.470779919	0
240331_at	5.483028051	0
236934_at	5.488241464	0.000266129
240758_at	5.505853056	8.78049E-05
236429_at	5.51260407	4.36162E-05
216550_x_at	5.518665672	0.010135825
240018_at	5.530308867	5.14403E-06
226625_at	5.540433592	8.46883E-05
227527_at	5.543064256	0.034436211
224862_at	5.547946417	0.009524126
209343_at	5.551179826	0.000575379
242343_x_at	5.562345277	0.020072202
244248_at	5.566229784	0.000115951
235049_at	5.569429609	1.059629263
223679_at	5.570733878	1.89125E-05
215698_at	5.590234331	0.003521676
201290_at	5.599359938	0.003094521
201622_at	5.615073168	1.058020969
228487_s_at	5.618637204	1.89484E-05
243525_at	5.61898866	2.10704E-05
216527_at	5.619239541	0.001249061
232288_at	5.623937537	9.91683E-05
210908_s_at	5.637451327	1.087592982
235697_at	5.641604661	1.46987E-05
1561686_at	5.665037692	0
230885_at	5.674984463	0.048861132
212995_x_at	5.683159382	1.042819088
229374_at	5.686883951	1.050706561
212451_at	5.694641211	0.00027496
227066_at	5.696653811	0.000800667
221886_at	5.705059241	0
205816_at	5.710120405	0
229065_at	5.716239015	1.036756208
238716_at	5.717198906	0.000223214
241741_at	5.720555938	0.000378825
218843_at	5.731052778	1.026193277
221497_x_at	5.736672449	0
222745_s_at	5.740658659	0.003341628

Comparison result of GeneMeta with Coloured (α, β)-*k* Feature Set and Generalised (α, β)-*k* Feature Set problem approach. (continued)

Probe ID	Z-Score	FDR
209751_s_at	5.749881918	1.012232869
1563881_at	5.750376094	0.003453436
204517_at	5.765395304	5.20021E-06
243704_at	5.767277142	1.022887753
1561158_at	5.778956269	0
210268_at	5.789799834	0
226501_at	5.796554902	0
51774_s_at	5.803489606	0.001469793
1557155_a_at	5.807392736	0
213079_at	5.813011687	1.067480044
223519_at	5.827080751	0.002599874
219671_at	5.83319271	1.01034151
238525_at	5.834095463	7.84593E-05
1560741_at	5.84463751	0
231530_s_at	5.867596815	1.009781081
1559949_at	5.870066354	0
243791_at	5.898284725	0
218520_at	5.901377951	1.069391308
221506_s_at	5.914332199	0.030377719
214623_at	5.918416036	0
203031_s_at	5.930060402	1.040236576
216308_x_at	5.938545826	1.022562163
239757_at	5.947638375	0
210686_x_at	5.949302745	1.79856E-05
225219_at	5.955414492	0.002173556
224613_s_at	5.963954698	1.013341003
206140_at	5.967120149	0.076878845
215553_x_at	5.972996447	8.22763E-05
225655_at	5.974234956	0
224568_x_at	5.982379423	0.000595423
201709_s_at	5.988139142	1.079430469
233816_at	5.999125252	0
223185_s_at	6.005317508	4.0032E-05
232173_at	6.017810846	1.037600935
1568763_s_at	6.02051436	0.000346863
212468_at	6.023380256	0.009934405
222457_s_at	6.037316255	0
203132_at	6.055336051	0.011851813
212372_at	6.057933038	1.008568949
242413_at	6.060320646	0
229497_at	6.065023946	0.000115553
239102_s_at	6.087113136	0.004024815
1561346_at	6.09149804	1.84502E-05
202135_s_at	6.094969624	1.019774433
213089_at	6.112404084	0
238560_at	6.112922708	0
218271_s_at	6.130725512	1.082087958
242240_at	6.13220344	0.00041716
238470_at	6.132377067	0
209250_at	6.139203584	0.056278552
226895_at	6.14509329	0.021290466
212209_at	6.145899323	0.001938037
222610_s_at	6.15663375	0
225636_at	6.158995605	0.000775681

Comparison result of GeneMeta with Coloured (α, β)-*k* Feature Set and Generalised (α, β)-*k* Feature Set problem approach. (continued)

Probe ID	Z-Score	FDR
212501_at	6.167513107	0.000378182
209177_at	6.167636348	1.030586315
226195_at	6.169860103	1.056356003
209169_at	6.186957339	0.057205576
214850_at	6.20117412	0
244045_at	6.204793909	0
227772_at	6.220867941	0.00186892
226143_at	6.234934498	0.003766197
1558792_x_at	6.235648044	0.000491803
240554_at	6.258508373	0.000171611
222047_s_at	6.258976204	0.020534194
227792_at	6.262290416	0
207132_x_at	6.268092837	1.08995752
201067_at	6.271802545	1.033058545
1566480_x_at	6.274005036	0
210840_s_at	6.281798126	0
224151_s_at	6.29226341	0.000630454
235730_at	6.298961985	6.40301E-05
240948_at	6.300467217	0
232311_at	6.302309112	0
227556_at	6.31649994	1.049141962
227435_at	6.318012111	6.38618E-05
225917_at	6.318801298	0.000431706
244027_at	6.321091058	0
236268_at	6.326451208	0.004617461
202858_at	6.337485449	1.063219051
234997_x_at	6.337648345	0.000415578
227798_at	6.342115822	0.000205128
217028_at	6.351776215	0
229309_at	6.359458455	1.46556E-05
238812_at	6.359547853	1.76523E-05
1562280_at	6.361574015	0
204731_at	6.365392205	3.63636E-05
207598_x_at	6.370016757	0
239096_at	6.370430357	0.000616798
233039_at	6.380100185	4.97018E-06
243561_at	6.388461785	0.00014826
229143_at	6.389637825	0.000382567
228287_at	6.414363797	0.000922144
226101_at	6.417173109	1.039823877
221952_x_at	6.426468145	1.065907999
225501_at	6.429018803	0.000245098
240602_at	6.430124694	0
201182_s_at	6.430418222	0
214246_x_at	6.431951889	0.053609711
209023_s_at	6.435383292	0.010730427
220269_at	6.446605799	1.038327249
242578_x_at	6.447960639	0
206547_s_at	6.454975645	1.05328858
218048_at	6.459486663	1.075543079
228839_s_at	6.459952983	0.027108672
205353_s_at	6.460711606	1.068158431
236901_at	6.48492403	0
231986_at	6.488463813	1.049062494

Comparison result of GeneMeta with Coloured (α, β)-k Feature Set and Generalised (α, β)-k Feature Set problem approach. (continued)

Probe ID	Z-Score	FDR
200976_s_at	6.49996442	1.085672224
208611_s_at	6.510504623	1.012871159
81811_at	6.516033892	0
1562416_at	6.519872353	0
214499_s_at	6.520852457	0
209163_at	6.521464803	1.043116154
222982_x_at	6.546949135	0
225864_at	6.55323323	0
219028_at	6.563874835	0
200954_at	6.573084803	1.010958701
213693_s_at	6.57545612	0
202121_s_at	6.579796134	1.05602085
224999_at	6.580726995	0
1553186_x_at	6.581978784	7.63359E-05
204720_s_at	6.587555774	1.005412471
224414_s_at	6.595570773	0
242319_at	6.608347072	0.000363326
207789_s_at	6.618937066	1.043394027
203509_at	6.63091557	1.081889715
238602_at	6.644252491	0.000222346
237018_at	6.651325804	5.58036E-06
231387_at	6.651911182	0
236139_at	6.654934596	0
212095_s_at	6.657204024	0
234163_at	6.658335208	3.64372E-05
236283_x_at	6.659951407	0.000531939
233437_at	6.661904727	1.042265631
203956_at	6.667875663	5.06428E-05
217969_at	6.67622006	1.014583488
1558831_x_at	6.678523371	0
226718_at	6.681107764	1.035324651
1568877_a_at	6.684792703	0.000760824
236752_at	6.686744413	0.000524157
224739_at	6.710246944	0
236484_at	6.718148554	0
209715_at	6.719420242	0
1556690_s_at	6.72370241	0.000142554
229966_at	6.725916507	0
36711_at	6.733142647	0
228662_at	6.734066349	1.80995E-05
222651_s_at	6.736141068	0.000558546
227802_at	6.743057193	0.001030279
236274_at	6.747084174	5.42425E-05
225571_at	6.748610799	0.000123077
212450_at	6.750060758	0.003973371
220072_at	6.751244121	0
222513_s_at	6.752337965	1.81818E-05
222043_at	6.754433593	0.000277038
241798_at	6.763800255	0
217726_at	6.764327932	1.023616291
218302_at	6.765576402	1.044580998
218247_s_at	6.771308355	0
206652_at	6.776446661	0
219032_x_at	6.783638978	1.037875862

Comparison result of GeneMeta with Coloured (α, β)-*k* Feature Set and Generalised (α, β)-*k* Feature Set problem approach. (continued)

Probe ID	Z-Score	FDR
215167_at	6.785760921	1.004598716
232653_at	6.789789862	0
219019_at	6.793336043	8.29016E-05
227959_at	6.798222326	0
222364_at	6.799893455	0.000333752
224972_at	6.806612822	1.035109174
213421_x_at	6.816418095	1.014395037
1552507_at	6.816724836	0
235567_at	6.817814388	5.54939E-06
227663_at	6.81913076	1.89934E-05
202927_at	6.831122031	1.011258217
1561657_at	6.847436733	8.34725E-05
212088_at	6.857599457	1.028163112
1553693_s_at	6.862955886	0
225003_at	6.862968379	1.028763327
201057_s_at	6.866666537	1.004506327
202201_at	6.875014372	1.010154632
212348_s_at	6.87569603	1.01815843
204559_s_at	6.878769674	1.00411848
225055_at	6.881805872	0
220316_at	6.883564036	0
202551_s_at	6.884501108	0
236265_at	6.888089407	1.72265E-05
202330_s_at	6.900405065	0
211386_at	6.907563204	8.17795E-05
206562_s_at	6.912711729	0
209609_s_at	6.916807008	1.010360202
202360_at	6.918189118	0.000155595
225870_s_at	6.923136208	1.038209194
200085_s_at	6.955880412	1.013397413
223024_at	6.962521232	1.016055814
227891_s_at	6.970581515	0
243329_at	6.981858604	6.41509E-05
216745_x_at	6.985096369	0
208972_s_at	6.985546912	1.058732861
1558695_at	6.995874643	0.000104232
236949_at	6.997895316	0
214436_at	7.001385236	1.043949456
213710_s_at	7.001686225	1.003786278
203431_s_at	7.003381506	1.037462956
225223_at	7.028490265	0.001071567
222160_at	7.037245146	0
212228_s_at	7.041923381	1.023673852
223673_at	7.049828356	0
216101_at	7.053990013	0
200754_x_at	7.056118901	1.066359071
238736_at	7.059103326	0.001191062
228007_at	7.071098256	0.00017625
238346_s_at	7.077804591	0
224878_at	7.080270266	1.067977528
205594_at	7.081012765	0.000229218
213545_x_at	7.085388427	1.005116407
213517_at	7.090491092	0.001364723
1560116_a_at	7.096783594	0

Comparison result of GeneMeta with Coloured (α, β)-*k* Feature Set and Generalised (α, β)-*k* Feature Set problem approach. (continued)

Probe ID	Z-Score	FDR
220615_s_at	7.097093048	1.01645294
203466_at	7.097977577	1.003417422
201836_s_at	7.101932372	1.006116321
1554593_s_at	7.107357953	1.003362116
214799_at	7.11380777	0.000337385
213457_at	7.121640028	9.88205E-05
49452_at	7.121725393	0
225050_at	7.129293826	1.069649391
224227_s_at	7.133324172	0
240005_at	7.137067609	1.00332525
43427_at	7.14063708	0
241727_x_at	7.147543691	0
203140_at	7.158684128	0
211930_at	7.160250664	0.000339277
228556_at	7.162511449	0.000312258
201500_s_at	7.165580199	1.019298606
226169_at	7.171158764	1.93517E-05
244803_at	7.173626806	0
239026_x_at	7.193864866	5.33049E-06
217077_s_at	7.196403316	1.031869214
234723_x_at	7.196885175	0
213195_at	7.203844535	1.02553828
243025_at	7.204649972	0
1316_at	7.206437799	0.002712444
207084_at	7.211090056	0
220071_x_at	7.216659093	0
236488_s_at	7.235285023	0.000335737
205751_at	7.237909858	1.002975152
209558_s_at	7.241192234	0
1569661_at	7.241365654	0
234989_at	7.252358934	0
236783_at	7.252818374	1.005282922
224666_at	7.267673266	1.023731419
225239_at	7.274153044	0
209553_at	7.282267777	1.002864646
230713_at	7.282746265	0.000874283
226112_at	7.297797331	1.85099E-05
212645_x_at	7.301899631	1.029305854
203814_s_at	7.306412775	1.002772575
200771_at	7.310228084	0
203606_at	7.310454209	1.0207465
235200_at	7.321682207	0.000566122
203297_s_at	7.323500254	0
230280_at	7.330657756	1.002643706
1569482_at	7.335361962	5.47945E-06
208704_x_at	7.357932184	1.067371572
215600_x_at	7.362835718	0
202534_x_at	7.37070556	0
217871_s_at	7.375414009	1.023060208
213744_at	7.379554708	1.022007223
241216_at	7.388748754	7.98334E-05
239179_at	7.399014656	0
217398_x_at	7.405697585	1.011351852
214375_at	7.415481009	0

Comparison result of GeneMeta with Coloured (α, β)-*k* Feature Set and Generalised (α, β)-*k* Feature Set problem approach. (continued)

Probe ID	Z-Score	FDR
235575_at	7.42534365	0
226544_x_at	7.42559231	0
204229_at	7.427911641	1.028008471
235964_x_at	7.431187075	0
243593_s_at	7.431931033	0
233216_at	7.437865381	0
239831_at	7.441175125	0
212384_at	7.443763911	7.60319E-05
241391_at	7.444784389	0
225750_at	7.450408102	1.010434976
214743_at	7.450625987	0
225923_at	7.454962617	1.048170371
201991_s_at	7.46453113	6.40543E-05
226447_at	7.46679468	0
229531_at	7.476367389	0
222380_s_at	7.477727914	0
224598_at	7.481254975	0
221488_s_at	7.497640969	1.030061865
212265_at	7.504599641	0
242712_x_at	7.509066522	5.30223E-06
227847_at	7.50946431	1.025634578
215543_s_at	7.517522711	1.036854626
210679_x_at	7.521690401	0
241893_at	7.525579166	0
212252_at	7.52621582	1.022428157
200980_s_at	7.531697351	1.084581795
218026_at	7.537674763	1.008419964
200693_at	7.543356697	1.002018238
230296_at	7.548581045	1.8315E-05
1570210_x_at	7.55032807	0
241786_at	7.556292498	0.000115625
49077_at	7.569756883	1.032999924
229315_at	7.573337012	0
231281_at	7.581578926	0
239629_at	7.58984523	0
242195_x_at	7.593239583	8.50629E-05
217937_s_at	7.594890879	0
217860_at	7.596244952	1.033760624
239035_at	7.602375438	0
230255_at	7.60554871	1.004765059
238851_at	7.614318236	1.83234E-05
201065_s_at	7.614330021	0.000335487
241954_at	7.623717651	0
200098_s_at	7.628755343	1.005264417
233702_x_at	7.633330227	0
202829_s_at	7.638977706	1.00179767
209289_at	7.643294844	0
236338_at	7.643304344	6.38858E-05
230923_at	7.664881166	1.023386114
233037_at	7.673544694	0
1569302_at	7.674930782	0
1559391_s_at	7.675495435	9.85692E-05
201305_x_at	7.677658497	0
215383_x_at	7.67866507	3.61882E-05

Comparison result of GeneMeta with Coloured (α, β)-k Feature Set and Generalised (α, β)-k Feature Set problem approach. (continued)

Probe ID	Z-Score	FDR
225172_at	7.692877778	0
208996_s_at	7.706702897	1.001632309
212581_x_at	7.706852268	1.007917466
225240_s_at	7.707260443	0
244610_x_at	7.718338934	0
203122_at	7.720357906	1.031460517
1559060_a_at	7.722012903	0
214823_at	7.728006956	1.027641126
212992_at	7.728513459	1.016509697
208835_s_at	7.729706801	6.4006E-05
221772_s_at	7.742855252	1.018006599
220942_x_at	7.754366188	1.058182699
213453_x_at	7.759315087	1.007062512
214594_x_at	7.775149484	0
63825_at	7.777325351	0
201509_at	7.778589025	1.005208909
221853_s_at	7.78251331	1.008736609
215385_at	7.785776518	9.8668E-06
200822_x_at	7.788761946	1.005171906
208942_s_at	7.791909436	0.000187104
234981_x_at	7.794118134	0
41220_at	7.797444406	0
212207_at	7.801738858	0
243827_at	7.810177918	0
202961_s_at	7.840971183	1.047516449
223480_s_at	7.84858545	1.001320108
223380_s_at	7.850192512	0
201527_at	7.859082985	1.001430274
212178_s_at	7.861258822	5.09424E-06
211779_x_at	7.863317718	1.020860983
201145_at	7.864936617	1.002588486
201415_at	7.867793102	1.002257295
233319_x_at	7.868557956	0
202868_s_at	7.87514207	1.039725274
202120_x_at	7.875535067	1.015696777
1565620_at	7.875709174	0
209991_x_at	7.876620441	1.021739537
214707_x_at	7.878273684	0
208859_s_at	7.879819353	0.000115337
237040_at	7.890997465	0
205559_s_at	7.896767105	0
242239_at	7.9101959	1.76211E-05
201704_at	7.912596334	1.007266825
204552_at	7.91621436	1.001191611
1555495_a_at	7.924610383	4.97265E-06
217761_at	7.92922402	0
233313_at	7.931444801	0
211752_s_at	7.944679338	1.028627879
201828_x_at	7.945746941	1.012608236
226999_at	7.94635031	0
201371_s_at	7.952483302	1.009258575
229537_at	7.988464908	1.003454295
205514_at	8.005478536	1.001099848
204270_at	8.017882096	1.7301E-05

Comparison result of GeneMeta with Coloured (α, β)-*k* Feature Set and Generalised (α, β)-*k* Feature Set problem approach. (continued)

Probe ID	Z-Score	FDR
227084_at	8.018648355	0
227944_at	8.024254438	1.005468002
1558678_s_at	8.030024628	0
204466_s_at	8.048725544	1.005356946
217939_s_at	8.060068677	1.004044638
229145_at	8.064112873	0
238017_at	8.074582601	1.002459663
217713_x_at	8.077415093	0
213808_at	8.087640633	1.023117705
243648_at	8.091782299	0
202144_s_at	8.101261475	1.010173316
201434_at	8.102115001	1.039705479
203146_s_at	8.107722688	1.015017192
226037_s_at	8.12008924	1.000953062
222154_s_at	8.122072241	1.004524804
219389_at	8.123194693	1.009762411
230790_x_at	8.130537736	0
232215_x_at	8.133388859	0
222395_s_at	8.134988701	1.74216E-05
225234_at	8.140344357	0
205787_x_at	8.14226647	0
219670_at	8.158906901	1.024288233
1568603_at	8.167929863	1.004247729
239497_at	8.172066909	0
217482_at	8.173675327	2.13311E-05
200732_s_at	8.183633302	1.011895277
215091_s_at	8.20145117	1.036560187
200639_s_at	8.204959767	1.000861342
204957_at	8.206397286	1.000843
200625_s_at	8.231590959	1.022638754
223035_s_at	8.233759808	1.010285439
212432_at	8.246554926	1.004303132
218330_s_at	8.247967035	0
224774_s_at	8.260191424	0
203068_at	8.269572914	0
219961_s_at	8.274622509	1.018538205
220966_x_at	8.279437228	1.032843823
207081_s_at	8.285307901	1.002404464
239333_x_at	8.289266008	0
242829_x_at	8.296484192	7.27273E-05
209268_at	8.311420757	1.007898865
201441_at	8.342452711	1.017797905
229353_s_at	8.344688895	0.000325451
216524_x_at	8.347370828	0
201570_at	8.348778496	1.003104107
204977_at	8.361279955	1.013322201
202178_at	8.395204788	1.015829024
212483_at	8.396235564	5.28541E-06
211464_x_at	8.399473031	0
225649_s_at	8.421851647	0
209001_s_at	8.423144969	1.001265034
222809_x_at	8.435604897	0
204786_s_at	8.459985224	0
212462_at	8.472153707	0

Comparison result of GeneMeta with Coloured (α, β)-k Feature Set and Generalised (α, β)-k Feature Set problem approach. (continued)

Probe ID	Z-Score	FDR
237768_x_at	8.493954499	1.46128E-05
244778_x_at	8.517478788	0
225356_at	8.519140286	5.54324E-06
208969_at	8.558919386	1.012965093
209225_x_at	8.576174302	0
212177_at	8.576245236	5.46747E-06
213015_at	8.584954399	0
205257_s_at	8.600178888	1.001944704
208936_x_at	8.615355212	1.003878534
224366_s_at	8.639988516	1.000494632
46665_at	8.648247822	0
212296_at	8.650870942	1.000439649
202974_at	8.677399052	1.010678066
203773_x_at	8.683847497	1.015281367
223037_at	8.691330009	1.000384672
225493_at	8.717693847	0
31874_at	8.733395333	5.03778E-06
226086_at	8.770740842	1.002146947
230392_at	8.786568765	0
218381_s_at	8.830730657	0
209251_x_at	8.831924399	1.011501704
220081_x_at	8.85151201	5.46448E-06
225092_at	8.85858329	1.001760919
212832_s_at	8.858598153	1.000293056
204663_at	8.859039873	1.002754163
201410_at	8.861787197	1.001044798
202070_s_at	8.877426476	1.013510253
211993_at	8.888995002	0
209583_s_at	8.903164312	1.000256415
228594_at	8.911605762	0
238558_at	8.912536059	0
203752_s_at	9.00400491	0
208549_x_at	9.006963137	0
209258_s_at	9.019977997	0
211921_x_at	9.026860592	0
212114_at	9.033005735	1.036697546
234969_s_at	9.069809342	5.66251E-06
217927_at	9.07738564	1.004617196
203909_at	9.098736125	1.008941602
1568612_at	9.102955539	1.001522098
218953_s_at	9.122404031	1.000128191
212491_s_at	9.133306073	1.000549622
230656_s_at	9.14454009	1.003767828
215021_s_at	9.159543923	1.004081558
200708_at	9.208228964	1.007936069
224689_at	9.217657499	1.001871182
223134_at	9.268611756	0
201400_at	9.332096358	1.000219776
223319_at	9.38252747	1.000806319
210976_s_at	9.398271664	1.001540465
202712_s_at	9.451228325	1.011145878
208732_at	9.470725787	1.012758461
224628_at	9.485526309	1.000036623
202160_at	9.525430489	0

Comparison result of GeneMeta with Coloured (α, β)-*k* Feature Set and Generalised (α, β)-*k* Feature Set problem approach. (continued)

Probe ID	Z-Score	FDR
200640_at	9.661444248	1.002901478
200039_s_at	9.746407641	1
221476_s_at	10.10649561	1.004968441

Probe ID is the list of probes that are present in the top listed probes in the GeneMeta. **Z-Score** indicates the z- score resulted from GeneMeta approach. **FDR** is the false discovery rate resulted from GeneMeta approach.

Sensitivity analysis result of AD datasets.

Table 8.14: **Sensitivity analysis result.**

Probe ID	BF-value	CABk	EC	HIP	MTG-PC	DS	Symbol
210976_s_at	5.75543E-23	X	X	X	X	4	PFKM
200039_s_at	8.0145E-23	X	X	X	X	4	PSMB2
211993_at	6.42503E-22	X	X	0	0	2	WNK1
221476_s_at	8.60884E-22	X	0	X	X	3	RPL15
211921_x_at	9.29541E-22	X	X	X	X	4	PTMA
46665_at	1.01939E-21	X	X	X	X	4	SEMA4C
238558_at	1.71324E-21	X	X	X	X	4	0
208549_x_at	3.57697E-21	X	X	X	X	4	PTMA
223319_at	2.26256E-19	X	X	X	X	4	GPHN
208732_at	7.52584E-19	X	X	X	X	4	RAB2A
213555_at	1.27531E-18	X	X	0	X	3	RWDD2A
203146_s_at	1.48232E-18	X	0	X	X	3	GABBR1
224567_x_at	1.91461E-18	X	X	X	X	4	MALAT1
212296_at	2.36066E-18	X	X	X	X	4	PSMD14
200708_at	3.18003E-18	X	X	X	X	4	GOT2
204786_s_at	3.29094E-18	X	X	X	X	4	IFNAR2
215543_s_at	3.43608E-18	X	X	X	X	4	LARGE
228027_at	3.76682E-18	X	X	0	0	2	GPRASP2
230392_at	5.22772E-18	X	0	X	X	3	0
1568604_a_at	8.01076E-18	X	X	0	0	2	CADPS
203517_at	1.28062E-17	X	0	0	0	1	MT
224689_at	3.39685E-17	X	0	X	X	3	MANBAL
236338_at	3.64697E-17	X	X	X	X	4	0
207170_s_at	4.42394E-17	X	X	0	0	2	LETMD1
215021_s_at	4.79697E-17	X	X	0	X	3	NR
223035_s_at	7.1786E-17	X	X	X	X	4	FARSB
227322_s_at	7.56397E-17	X	0	0	0	1	BCCIP
225649_s_at	9.2195E-17	X	X	X	X	4	STK35
211615_s_at	1.45681E-16	X	X	0	0	2	LRPPRC
218138_at	1.5198E-16	X	X	0	0	2	MKKS
223037_at	1.93984E-16	X	X	X	X	4	PDZD11
1553274_a_at	2.68246E-16	X	X	0	0	2	SNRNP48
201065_s_at	2.72913E-16	X	0	X	X	3	0
227404_s_at	3.01186E-16	X	0	X	X	3	EGR1
209583_s_at	3.23885E-16	X	0	X	X	3	CD200
226377_at	3.53728E-16	X	X	0	0	2	0
238851_at	3.5945E-16	X	X	X	X	4	ANKRD13A
203068_at	3.74244E-16	X	X	X	X	4	KLHL21
204270_at	3.87413E-16	X	X	X	X	4	SKI

Sensitivity analysis result. (continued)

Probe ID	BF-value	CABk	EC	HIP	MTG-PC	DS	Symbol
209001_s_at	4.09753E-16	X	X	X	X	4	ANAPC13
203264_s_at	4.21967E-16	X	X	X	0	3	ARHGEF9
218953_s_at	5.31899E-16	X	X	X	X	4	PCYO
225172_at	8.38837E-16	X	X	X	X	4	CRAMP1L
212450_at	8.61555E-16	X	0	X	X	3	SECISBP2L
225234_at	8.95423E-16	X	X	X	X	4	CBL
201837_s_at	9.2725E-16	X	0	0	0	1	SUPT7L
211930_at	1.03861E-15	X	0	X	X	3	HNRNPA3
223107_s_at	1.1542E-15	X	X	0	0	2	ZCCHC17
218671_s_at	1.34339E-15	X	X	0	0	2	ATPIF1
201002_s_at	1.53325E-15	X	0	0	0	1	0
200640_at	1.64896E-15	X	X	X	X	4	YWHAZ
200980_s_at	1.69765E-15	X	X	X	X	4	PDHA1
212832_s_at	1.73144E-15	X	X	X	X	4	CKAP5
239333_x_at	2.00702E-15	X	X	X	X	4	0
205787_x_at	2.34926E-15	X	X	X	X	4	ZC3H11A
209163_at	2.6735E-15	X	0	X	X	3	CYB561
213015_at	2.97327E-15	X	0	X	X	3	BB
202144_s_at	3.4052E-15	X	X	X	X	4	ADSL
227944_at	3.55744E-15	X	X	0	X	3	PTPN3
208722_s_at	3.98205E-15	X	0	0	0	1	ANAPC5
224879_at	4.06852E-15	X	X	0	0	2	TMEM261
1555495_a_at	4.74325E-15	X	0	X	X	3	CWC27
217761_at	5.75884E-15	X	0	X	X	3	ADI1
50221_at	5.94941E-15	X	X	0	0	2	TFEB
219670_at	6.09459E-15	X	X	X	0	3	BEND5
205047_s_at	6.21392E-15	X	X	0	0	2	ASNS
201758_at	8.28856E-15	X	0	0	0	1	TSG101
222809_x_at	8.54267E-15	X	0	X	X	3	CCDC85C
1569302_at	8.54621E-15	X	X	X	X	4	CEP295
208969_at	8.62671E-15	X	X	X	X	4	NDUFA9
201305_x_at	8.9563E-15	X	X	X	X	4	ANP32B
212978_at	9.24151E-15	X	0	0	0	1	LRRC8B
208616_s_at	9.30663E-15	X	0	0	0	1	PTP4A2
232940_s_at	9.36944E-15	X	0	0	0	1	KMT2C
213457_at	1.21659E-14	X	0	X	X	3	MFHAS1
212122_at	1.71139E-14	X	0	0	0	1	RHOQ
210640_s_at	1.76542E-14	X	0	0	0	1	GPER1
241216_at	1.96718E-14	X	0	X	X	3	0
205514_at	2.79443E-14	X	X	X	X	4	ZNF415
235225_at	3.32794E-14	X	X	0	0	2	SCN2B
212645_x_at	4.24423E-14	X	0	X	X	3	BRE
215529_x_at	4.28368E-14	X	0	0	0	1	DIP2A
217941_s_at	4.76407E-14	X	0	0	0	1	ERBB2IP
36907_at	5.03004E-14	X	X	0	0	2	MVK
233313_at	5.16759E-14	X	X	X	X	4	0
223461_at	5.3378E-14	X	X	0	X	3	TBC1D7
211995_x_at	5.86758E-14	X	X	0	0	2	ACTG1
1558678_s_at	6.74337E-14	X	X	X	X	4	MALAT1
224462_s_at	6.94615E-14	X	X	0	0	2	CHCHD6
229793_at	7.69084E-14	X	0	0	0	1	ASAH2B
218788_s_at	8.12646E-14	X	X	0	0	2	SMYD3
203466_at	9.07151E-14	X	X	X	X	4	MPV17
201391_at	9.18014E-14	X	0	X	X	3	TRAP1
212491_s_at	9.48778E-14	X	0	X	X	3	DNAJC8

Sensitivity analysis result. (continued)

Probe ID	BF-value	CABk	EC	HIP	MTG-PC	DS	Symbol
201676_x_at	9.56967E-14	X	X	0	0	2	PSMA1
209953_s_at	9.82056E-14	X	0	0	0	1	CDC37
215383_x_at	1.02069E-13	X	X	X	X	4	SPG21
224413_s_at	1.04044E-13	X	0	0	0	1	TM2D2
208713_at	1.05483E-13	X	X	0	0	2	HNRNPUL1
217964_at	1.14536E-13	X	0	0	0	1	TTC19
220690_s_at	1.24708E-13	X	0	0	0	1	DHRS7B
209153_s_at	1.2969E-13	X	X	0	0	2	TCF3
201924_at	1.36874E-13	X	0	0	0	1	AFF1
233216_at	1.38778E-13	X	X	X	X	4	ZDHHC21
200732_s_at	1.56578E-13	X	X	X	X	4	PTP4A1
226037_s_at	1.65114E-13	X	0	X	X	3	TAF9B
210425_x_at	1.7701E-13	X	X	X	0	3	GOLGA8B
225917_at	2.46802E-13	X	0	X	X	3	0
235575_at	2.57672E-13	X	0	X	X	3	0
239035_at	2.62182E-13	X	X	X	X	4	MTHFR
41220_at	3.15069E-13	X	0	X	X	3	0
205531_s_at	3.25655E-13	X	0	0	0	1	GLS2
201371_s_at	3.30293E-13	X	X	X	X	4	CUL3
221974_at	3.65658E-13	X	0	0	0	1	IPW
222364_at	4.08035E-13	X	0	X	X	3	SLC44A1
225219_at	4.12394E-13	X	0	X	X	3	SMAD5
202364_at	4.18667E-13	X	0	0	0	1	M
235567_at	4.29608E-13	X	0	X	X	3	RORA
211566_x_at	4.39372E-13	X	0	0	0	1	BRE
218824_at	4.5724E-13	X	X	0	0	2	PNMAL1
234997_x_at	4.75626E-13	X	0	X	X	3	0
226022_at	4.97814E-13	X	0	0	0	1	SASH1
1568877_a_at	5.56958E-13	X	0	X	X	3	ACBD5
202534_x_at	5.60754E-13	X	X	0	X	3	DHFR
215385_at	5.61987E-13	X	X	X	X	4	0
213122_at	5.79538E-13	X	0	0	0	1	TSPYL5
238017_at	7.53603E-13	X	X	X	X	4	SDR16C5
224366_s_at	7.54373E-13	X	X	X	X	4	REPS1
39891_at	7.58618E-13	X	X	X	0	3	ZNF710
204228_at	7.72855E-13	X	X	0	X	3	PPIH
200639_s_at	8.16223E-13	X	0	X	X	3	YWHAZ
36711_at	1.15572E-12	X	X	X	X	4	MAFF
214436_at	1.19012E-12	X	0	X	X	3	FB
224509_s_at	1.26771E-12	X	X	0	0	2	RTN4IP1
1553693_s_at	1.28585E-12	X	0	X	X	3	CBR4
239831_at	1.28639E-12	X	X	X	X	4	TMEM106C
229531_at	1.39887E-12	X	X	X	X	4	0
223024_at	1.48719E-12	X	X	X	X	4	AP1M1
203769_s_at	1.66563E-12	X	0	0	0	1	STS
214375_at	1.82121E-12	X	X	X	X	4	0
204159_at	1.86486E-12	X	0	0	0	1	CDKN2C
223804_s_at	1.88854E-12	X	0	0	0	1	THUMPD3
202551_s_at	1.93595E-12	X	0	X	X	3	CRIM1
224739_at	2.50802E-12	X	0	X	X	3	PIM3
49077_at	2.67667E-12	X	0	X	X	3	PPME1
214728_x_at	2.97624E-12	X	X	0	0	2	SMARCA4
229687_s_at	3.3742E-12	X	0	X	X	3	LOC100287017
203296_s_at	3.5394E-12	X	0	0	0	1	ATP1A2
224780_at	3.59675E-12	X	0	0	0	1	RBM17

Sensitivity analysis result. (continued)

Probe ID	BF-value	CABk	EC	HIP	MTG-PC	DS	Symbol
206551_x_at	3.65061E-12	X	0	0	0	1	KLHL24
1316_at	3.76938E-12	X	X	X	X	4	THRA
213151_s_at	4.15086E-12	X	0	0	0	1	0
229537_at	4.59659E-12	X	0	X	X	3	LMO4
227959_at	4.87122E-12	X	0	X	X	3	0
230559_x_at	5.36671E-12	X	0	0	0	1	FGD4
203017_s_at	5.73159E-12	X	0	0	0	1	SS
207084_at	6.01962E-12	X	0	X	X	3	POU3F2
227891_s_at	6.05355E-12	X	X	X	X	4	TAF15
201592_at	6.6369E-12	X	X	0	0	2	EIF3H
1559003_a_at	7.13127E-12	X	0	X	0	2	CCDC163P
225368_at	7.19436E-12	X	0	0	0	1	HIPK2
202041_s_at	7.36355E-12	X	X	0	0	2	FIBP
236484_at	7.47077E-12	X	X	X	X	4	0
201550_x_at	7.71711E-12	X	0	0	0	1	ACTG1
202330_s_at	7.93061E-12	X	0	X	X	3	UNG
208704_x_at	7.93416E-12	X	X	X	X	4	APLP2
201938_at	8.17583E-12	X	0	X	0	2	CDK2AP1
209332_s_at	8.27793E-12	X	0	0	0	1	MA
235119_at	1.35316E-11	X	0	0	0	1	TAF3
201987_at	1.42594E-11	X	X	0	X	3	MED13
206562_s_at	1.45427E-11	X	X	X	X	4	CSNK1A1
205194_at	1.59129E-11	X	X	0	0	2	PSPH
223519_at	1.59162E-11	X	0	X	X	3	ZAK
212501_at	1.74252E-11	X	0	X	X	3	CEBPB
203620_s_at	2.11123E-11	X	0	0	0	1	FCHSD2
238009_at	2.17506E-11	X	X	0	X	3	0
230630_at	2.46508E-11	X	0	0	0	1	AK4
224999_at	2.59606E-11	X	0	0	X	2	0
213315_x_at	2.67795E-11	X	X	0	0	2	C
212114_at	2.86393E-11	X	X	X	X	4	AT
214779_s_at	3.01777E-11	X	0	0	0	1	SGSM3
233039_at	3.21958E-11	X	0	X	X	3	0
225788_at	3.93164E-11	X	X	0	0	2	RRP36
212348_s_at	3.94526E-11	X	0	X	X	3	KDM1A
212244_at	3.95305E-11	X	0	0	0	1	0
203336_s_at	4.12494E-11	X	X	X	0	3	ITGB1BP1
225117_at	4.43729E-11	X	0	0	0	1	KANSL1
223112_s_at	4.46087E-11	X	0	0	0	1	NDUFB10
81811_at	4.5774E-11	X	0	X	X	3	0
209411_s_at	4.67168E-11	X	0	0	0	1	GGA3
1558546_at	4.85019E-11	X	X	0	0	2	DNASE1
225356_at	5.33936E-11	X	X	X	X	4	0
200811_at	5.34891E-11	X	0	0	0	1	CIRBP
218133_s_at	5.4675E-11	X	X	0	0	2	NIF3L1
225387_at	5.60986E-11	X	0	0	0	1	TSPAN5
208936_x_at	5.79999E-11	X	0	X	X	3	LGALS8
205559_s_at	5.98857E-11	X	X	X	X	4	PCSK5
210381_s_at	6.34159E-11	X	0	0	0	1	CCKBR
218059_at	7.3521E-11	X	0	0	0	1	ZNF706
200875_s_at	8.13918E-11	X	0	0	0	1	NOP56
202829_s_at	8.57476E-11	X	0	X	X	3	VAMP7
210250_x_at	9.42884E-11	X	0	0	0	1	ADSL
222513_s_at	9.51247E-11	X	0	X	X	3	SORBS1
210357_s_at	9.65038E-11	X	0	0	0	1	SMO

Sensitivity analysis result. (continued)

Probe ID	BF-value	CABk	EC	HIP	MTG-PC	DS	Symbol
200720_s_at	9.75566E-11	X	0	0	X	2	ACTR1A
200954_at	1.17513E-10	X	0	X	X	3	ATP6V0C
213388_at	1.18739E-10	X	0	0	0	1	PDE4DIP
222154_s_at	1.24652E-10	X	0	X	X	3	SPATS2L
202176_at	1.37843E-10	X	0	0	0	1	ERCC3
1554593_s_at	1.41338E-10	X	0	X	X	3	SLC1A6
217840_at	1.43031E-10	X	0	0	0	1	DD
242413_at	1.52012E-10	X	0	X	X	3	0
202838_at	1.60578E-10	X	X	0	0	2	FUCA1
1559419_at	1.8843E-10	X	X	0	0	2	CACNB2
201463_s_at	1.88586E-10	X	X	0	0	2	TALDO1
230213_at	2.0528E-10	X	0	0	0	1	C19orf43
235805_at	2.06259E-10	X	0	0	0	1	0
212132_at	2.24443E-10	X	0	X	0	2	LSM14A
207949_s_at	2.47096E-10	X	0	0	0	1	ICA1
202432_at	2.50529E-10	X	0	0	0	1	PPP3CB
212833_at	2.52099E-10	X	0	0	0	1	SLC25A46
222423_at	2.83211E-10	X	X	0	0	2	NDFIP1
220155_s_at	2.83259E-10	X	0	0	0	1	BRD9
242139_s_at	2.88111E-10	X	0	0	0	1	0
241798_at	2.91994E-10	X	X	X	X	4	0
226443_at	2.92526E-10	X	X	0	0	2	FAM122A
200068_s_at	2.96121E-10	X	0	0	0	1	CAN
224925_at	3.17744E-10	X	0	0	0	1	PRE
223062_s_at	3.2244E-10	X	0	0	0	1	PSAT1
219640_at	3.37248E-10	X	X	0	0	2	CLDN15
233437_at	3.44688E-10	X	0	X	X	3	GABRA4
1553313_s_at	3.4679E-10	X	X	X	0	3	SLC5A3
201323_at	3.83672E-10	X	0	0	0	1	EBNA1BP2
1561158_at	4.07278E-10	X	0	X	X	3	0
222982_x_at	4.07695E-10	X	0	X	X	3	SLC38A2
211955_at	4.39337E-10	X	0	0	0	1	IPO5
226232_at	5.25197E-10	X	0	0	0	1	GDF11
238919_at	5.36306E-10	X	0	X	X	3	0
238560_at	5.44738E-10	X	0	X	X	3	CALCOCO2
1558792_x_at	5.46963E-10	X	0	X	X	3	AP2A1
37950_at	5.58783E-10	X	0	0	0	1	PREP
212360_at	6.06314E-10	X	0	0	0	1	AMPD2
229966_at	6.07779E-10	X	0	X	X	3	EWSR1
223145_s_at	6.17941E-10	X	0	0	0	1	AKIRIN2
226770_at	6.46985E-10	X	0	0	0	1	MAGI3
225534_at	6.61149E-10	X	0	0	0	1	SMIM19
217882_at	6.61925E-10	X	X	0	0	2	EMC3
224598_at	6.63402E-10	X	X	X	X	4	MGAT4B
208986_at	6.65948E-10	X	0	0	0	1	TCF12
211983_x_at	6.78059E-10	X	0	0	0	1	ACTG1
203405_at	7.17132E-10	X	X	0	0	2	PSMG1
1566638_at	7.45466E-10	X	0	0	0	1	0
210987_x_at	8.07437E-10	X	X	0	X	3	TPM1
219564_at	8.14528E-10	X	0	0	0	1	KCNJ16
201067_at	8.27041E-10	X	0	X	X	3	PSMC2
212726_at	8.43065E-10	X	X	0	0	2	PHF2
227062_at	8.60368E-10	X	X	0	0	2	0
210268_at	8.66861E-10	X	0	X	X	3	NF
212119_at	8.98326E-10	X	X	0	0	2	RHOQ

Sensitivity analysis result. (continued)

Probe ID	BF-value	CABk	EC	HIP	MTG-PC	DS	Symbol
242319_at	1.05654E-09	X	X	0	X	3	DGKG
223350_x_at	1.10661E-09	X	0	0	0	1	LIN7C
210027_s_at	1.13017E-09	X	0	0	0	1	APE
218763_at	1.14087E-09	X	X	0	0	2	ST
212417_at	1.26696E-09	X	0	0	0	1	SCAMP1
204842_x_at	1.27002E-09	X	X	0	0	2	PRKAR2A
240948_at	1.31401E-09	X	0	X	X	3	0
212165_at	1.33304E-09	X	X	0	0	2	0
236901_at	1.34707E-09	X	X	X	0	3	0
200067_x_at	1.37775E-09	X	0	0	0	1	SN
213405_at	1.40862E-09	X	0	0	0	1	RAB22A
238029_s_at	1.45546E-09	X	0	0	0	1	SLC16A14
208777_s_at	1.58217E-09	X	X	0	0	2	PSMD11
215167_at	1.64787E-09	X	0	X	X	3	MED14
205773_at	1.71253E-09	X	0	0	0	1	CPEB3
222445_at	1.86207E-09	X	0	0	0	1	SLC39A9
220072_at	1.98837E-09	X	0	X	X	3	CSPP1
242819_at	2.04156E-09	X	0	0	0	1	0
207132_x_at	2.22857E-09	X	X	X	X	4	PFDN5
224895_at	2.88843E-09	X	0	0	0	1	YAP1
231817_at	3.10455E-09	X	0	0	0	1	USP53
229676_at	3.44807E-09	X	0	0	0	1	MTPAP
227195_at	3.53096E-09	X	0	0	0	1	ZNF503
208003_s_at	3.55636E-09	X	X	0	0	2	NFAT5
212988_x_at	3.69731E-09	X	0	0	0	1	ACTG1
211758_x_at	3.80467E-09	X	0	0	0	1	T
244327_at	3.87668E-09	X	0	0	0	1	0
238567_at	3.92207E-09	X	0	0	0	1	SGPP2
1560145_at	4.33043E-09	X	0	0	0	1	MKLN1
223272_s_at	4.50628E-09	X	0	0	0	1	NTPCR
228763_at	4.65394E-09	X	X	0	0	2	MDP1
1557293_at	4.67857E-09	X	X	X	X	4	LINC00969
201415_at	4.9436E-09	X	0	X	X	3	GSS
213214_x_at	5.57741E-09	X	0	0	0	1	ACTG1
228662_at	5.9894E-09	X	X	X	X	4	SOCS7
217726_at	6.1632E-09	X	X	X	X	4	COPZ1
216527_at	6.21055E-09	X	X	X	X	4	0
215907_at	6.91933E-09	X	0	0	0	1	0
217028_at	8.70814E-09	X	X	X	X	4	0
213794_s_at	8.79252E-09	X	0	0	0	1	NGDN
243791_at	8.85126E-09	X	0	X	X	3	0
221699_s_at	9.52948E-09	X	0	0	0	1	DD
224217_s_at	9.83049E-09	X	0	0	0	1	FAF1
204300_at	9.88402E-09	X	X	0	0	2	GATB
243121_x_at	1.19531E-08	X	0	0	0	1	0
232195_at	1.2462E-08	X	0	0	0	1	GPR158
214314_s_at	1.34706E-08	X	0	0	0	1	EIF5B
212282_at	1.349E-08	X	0	0	0	1	TMEM97
207435_s_at	1.39166E-08	X	0	0	0	1	SRRM2
241998_at	1.57036E-08	X	0	0	0	1	C2orf80
238525_at	1.67443E-08	X	0	X	X	3	DD
1552536_at	1.79682E-08	X	0	0	0	1	VTI1A
225313_at	1.80563E-08	X	0	0	0	1	FAM217B
226365_at	1.82054E-08	X	0	0	0	1	0
239474_at	1.82368E-08	X	0	0	0	1	0

Sensitivity analysis result. (continued)

Probe ID	BF-value	CABk	EC	HIP	MTG-PC	DS	Symbol
1561686_at	1.91357E-08	X	0	X	X	3	0
225050_at	2.11028E-08	X	0	X	X	3	ZNF512
238701_x_at	2.14627E-08	X	0	0	0	1	COLCA1
200620_at	2.29685E-08	X	0	0	0	1	TMEM59
210524_x_at	2.44101E-08	X	0	0	0	1	0
202138_x_at	2.62204E-08	X	0	X	X	3	AIMP2
230839_at	2.80898E-08	X	0	0	0	1	PRMT8
203440_at	3.06584E-08	X	0	0	0	1	CDH2
219387_at	3.09614E-08	X	0	X	X	3	CCDC88A
243704_at	3.16056E-08	X	0	X	X	3	0
226879_at	3.28571E-08	X	0	0	0	1	HVCN1
230886_at	3.60712E-08	X	0	0	0	1	0
209009_at	3.80468E-08	X	0	0	0	1	ESD
203485_at	4.12283E-08	X	0	X	X	3	RTN1
228028_at	4.29047E-08	X	0	0	0	1	GAREML
240602_at	4.51452E-08	X	X	X	X	4	HBS1L
1562898_at	4.82988E-08	X	X	X	X	4	0
218381_s_at	4.89542E-08	X	X	X	X	4	U2AF2
226009_at	5.0031E-08	X	0	X	X	3	DPCD
229448_at	5.61165E-08	X	0	0	0	1	CERS1
201324_at	6.06392E-08	X	0	X	X	3	EMP1
224726_at	6.33086E-08	X	0	0	0	1	MIB1
201854_s_at	6.65121E-08	X	0	X	X	3	ATMIN
209459_s_at	6.86223E-08	X	X	X	X	4	ABAT
243184_at	7.31882E-08	X	0	0	0	1	0
212178_s_at	7.32248E-08	X	0	X	X	3	0
222796_at	7.46625E-08	X	0	0	0	1	PTCD1
1555318_at	7.51059E-08	X	X	0	0	2	HIF3A
203152_at	7.56265E-08	X	0	0	0	1	MRPL40
222651_s_at	8.59427E-08	X	0	X	X	3	TRPS1
210686_x_at	8.9914E-08	X	0	X	X	3	SLC25A16
226017_at	9.24126E-08	X	0	0	0	1	CMTM7
210908_s_at	9.53429E-08	X	X	X	X	4	PFDN5
219028_at	9.53911E-08	X	0	X	X	3	HIPK2
228776_at	1.05679E-07	X	0	0	0	1	GJC1
233816_at	1.07467E-07	X	0	X	X	3	0
219204_s_at	1.20509E-07	X	0	X	X	3	SRR
201569_s_at	1.2298E-07	X	0	0	0	1	SAMM50
227484_at	1.33371E-07	X	0	0	0	1	SRGAP1
212356_at	1.38294E-07	X	0	0	0	1	KHNYN
219343_at	1.45446E-07	X	0	0	0	1	CDC37L1
206039_at	1.45547E-07	X	0	0	0	1	RAB33A
211970_x_at	1.53905E-07	X	0	0	0	1	ACTG1
209250_at	1.55713E-07	X	X	X	X	4	DEGS1
218543_s_at	1.60406E-07	X	0	X	X	3	PARP12
200976_s_at	1.65951E-07	X	X	X	X	4	TA
1554079_at	1.66597E-07	X	0	0	0	1	GALNT18
1559044_at	1.67681E-07	X	X	0	0	2	E
209362_at	1.68491E-07	X	X	0	0	2	MED21
214087_s_at	1.92548E-07	X	0	0	0	1	MYBPC1
224585_x_at	2.00242E-07	X	0	0	0	1	ACTG1
201529_s_at	2.21115E-07	X	0	0	0	1	RPA1
227064_at	2.26139E-07	X	0	0	0	1	ANKRD40
229798_s_at	2.29696E-07	X	0	0	0	1	0
215714_s_at	2.34477E-07	X	0	0	0	1	SMARCA4

Sensitivity analysis result. (continued)

Probe ID	BF-value	CABk	EC	HIP	MTG-PC	DS	Symbol
231311_at	2.35681E-07	X	0	0	0	1	0
221432_s_at	2.50574E-07	X	0	0	0	1	SLC25A28
218140_x_at	2.64149E-07	X	0	0	0	1	SRPRB
208668_x_at	2.73241E-07	X	0	0	0	1	HMGN2
221236_s_at	2.77537E-07	X	X	0	0	2	STMN4
227904_at	2.85068E-07	X	0	0	0	1	AZI2
235697_at	2.85414E-07	X	0	X	X	3	0
232997_at	2.98708E-07	X	0	0	0	1	0
226933_s_at	3.19041E-07	X	0	0	0	1	LOC100287917
1552257_a_at	3.3513E-07	X	0	0	0	1	TTLL12
219440_at	3.3802E-07	X	0	0	0	1	RAI2
222759_at	3.47279E-07	X	0	0	0	1	SUV420H1
201939_at	3.4939E-07	X	0	0	0	1	PLK2
200783_s_at	3.56603E-07	X	0	0	0	1	STMN1
223250_at	3.80903E-07	X	0	0	0	1	KLHL7
227260_at	3.8541E-07	X	0	0	0	1	ANKRD10
218263_s_at	4.34491E-07	X	0	0	0	1	ZBED5
238545_at	4.42113E-07	X	X	0	0	2	BRD7
1552506_at	4.44602E-07	X	0	0	0	1	CRB2
207826_s_at	4.45316E-07	X	0	0	0	1	ID3
221886_at	4.49844E-07	X	0	X	X	3	DENND2A
236204_at	4.52073E-07	X	0	0	0	1	COPS8
1554542_at	4.71761E-07	X	0	0	0	1	SLC25A48
225731_at	5.13647E-07	X	0	0	0	1	ANKRD50
226750_at	5.17386E-07	X	0	0	0	1	LARP1B
224932_at	5.40223E-07	X	0	0	0	1	CHCHD10
229065_at	5.89102E-07	X	0	X	X	3	SLC35F3
235706_at	5.896E-07	X	0	0	0	1	CPM
224983_at	7.11435E-07	X	0	0	0	1	SCARB2
219359_at	7.14921E-07	X	0	0	0	1	ATHL1
240857_at	7.60138E-07	X	0	0	0	1	DNAH9
232653_at	7.91171E-07	X	0	X	X	3	0
218152_at	8.3171E-07	X	X	0	0	2	HMG20A
206258_at	1.009E-06	X	0	0	0	1	ST8SIA5
203565_s_at	1.10494E-06	X	0	0	0	1	MNAT1
221607_x_at	1.10617E-06	X	0	0	0	1	ACTG1
238786_at	1.12711E-06	X	0	0	0	1	0
203632_s_at	1.14621E-06	X	0	0	0	1	GPRC5B
201057_s_at	1.15045E-06	X	0	X	X	3	GOLGB1
212281_s_at	1.40977E-06	X	X	0	X	3	TMEM97
210048_at	1.69706E-06	X	0	0	0	1	NAPG
217367_s_at	1.85309E-06	X	0	0	0	1	ZH
208717_at	2.05127E-06	X	0	0	0	1	O
202936_s_at	2.20967E-06	X	0	0	0	1	SO
200023_s_at	2.28745E-06	X	0	0	0	1	EIF3F
220415_at	2.28766E-06	X	X	0	0	2	TNNI3K
243667_at	2.49394E-06	X	0	0	0	1	0
241696_at	2.51751E-06	X	0	0	0	1	0
220775_s_at	2.59087E-06	X	0	0	0	1	UEVLD
235049_at	3.07386E-06	X	0	X	X	3	ADCY1
1558733_at	3.21851E-06	X	X	0	X	3	SOAT1
244661_at	3.44195E-06	X	X	0	0	2	LOC100292959
213789_at	3.48275E-06	X	0	0	0	1	0
242888_at	4.07521E-06	X	0	0	0	1	SLC16A6
230748_at	4.16811E-06	X	0	0	0	1	SPAG7

Sensitivity analysis result. (continued)

Probe ID	BF-value	CABk	EC	HIP	MTG-PC	DS	Symbol
200053_at	4.29923E-06	X	X	X	X	4	SNW1
222183_x_at	4.57071E-06	X	0	0	0	1	PRPF4
209161_at	4.58522E-06	X	0	0	0	1	TMCC1
227112_at	4.69889E-06	X	0	0	0	1	PDGFA
229830_at	4.70455E-06	X	0	0	0	1	MRPS30
218398_at	5.38592E-06	X	0	0	0	1	RBM12
212170_at	5.39143E-06	X	0	0	0	1	DDR2
225442_at	5.54858E-06	X	0	0	0	1	MAPK14
202530_at	5.59367E-06	X	0	0	0	1	SMG5
34868_at	5.64306E-06	X	0	0	0	1	FO
206015_s_at	5.70619E-06	X	0	0	0	1	DD
220890_s_at	5.75903E-06	X	0	0	0	1	CSK
202329_at	6.09109E-06	X	X	0	0	2	MRPL13
218049_s_at	6.3007E-06	X	0	0	0	1	AGAP3
239026_x_at	6.47617E-06	X	X	X	X	4	LAMB1
211651_s_at	6.67471E-06	X	0	0	0	1	OIP5-AS1
225225_at	7.17672E-06	X	0	0	0	1	ZNF32
209538_at	7.65549E-06	X	0	0	0	1	ELMO1
204513_s_at	7.69119E-06	X	0	0	0	1	F
201636_at	9.94309E-06	X	0	X	X	3	NDN
209550_at	1.03917E-05	X	0	0	0	1	0
238273_at	1.05152E-05	X	0	0	0	1	NDC1
234672_s_at	1.07005E-05	X	0	0	0	1	0
1569192_at	1.09992E-05	X	0	0	0	1	ANK3
209442_x_at	1.12088E-05	X	0	0	0	1	NKRF
205004_at	1.21452E-05	X	0	0	0	1	ARHGEF18
213039_at	1.31094E-05	X	0	X	X	3	CDHR3
231582_at	1.34611E-05	X	0	0	0	1	MAP1S
218522_s_at	1.37165E-05	X	0	X	X	3	ESAM
225369_at	1.37403E-05	X	0	0	0	1	NAMPT
217738_at	1.4944E-05	X	0	0	0	1	ZNF449
228968_at	1.58111E-05	X	0	0	0	1	0
1557155_a_at	2.10597E-05	X	X	X	X	4	0
1555653_at	2.15484E-05	X	X	0	0	2	SCAI
228174_at	2.19554E-05	X	0	0	0	1	SHANK2
213308_at	2.43799E-05	X	0	0	0	1	JADE1
235024_at	2.49423E-05	X	0	0	0	1	O
219133_at	2.95615E-05	X	0	0	0	1	EIF4A1
201530_x_at	3.14645E-05	X	0	0	0	1	FKSG49
224284_x_at	3.19687E-05	X	X	0	0	2	0
201736_s_at	3.60379E-05	X	0	0	0	1	MEG3
1552507_at	3.8277E-05	X	0	X	X	3	NEO1
1553186_x_at	3.8377E-05	X	0	X	X	3	ACSL6
1556690_s_at	3.8477E-05	X	0	X	X	3	TRMT5
1557895_at	3.8577E-05	X	0	X	X	3	PYGB
222328_x_at	3.86738E-05	X	0	0	0	1	RND1
1558279_a_at	3.8677E-05	X	0	X	X	3	PL
1558695_at	3.8777E-05	X	0	X	X	3	0
1558831_x_at	3.88771E-05	X	0	X	X	3	0
1559949_at	3.89771E-05	X	0	X	X	3	GAB2
1560116_a_at	3.90771E-05	X	0	X	X	3	EPHB6
1560689_s_at	3.91771E-05	X	0	X	X	3	AN
1560741_at	3.92771E-05	X	0	X	X	3	VEGFA
1562416_at	3.93771E-05	X	0	X	X	3	0
1563781_at	3.94771E-05	X	0	X	X	3	0

Sensitivity analysis result. (continued)

Probe ID	BF-value	CABk	EC	HIP	MTG-PC	DS	Symbol
1565620_at	3.95771E-05	X	0	X	X	3	TSN
1568603_at	3.97771E-05	X	0	X	X	3	MTO1
1568612_at	3.98771E-05	X	0	X	X	3	HBS1L
1568763_s_at	3.99771E-05	X	0	X	X	3	0
1569482_at	4.00771E-05	X	0	X	X	3	ACSL3
1570210_x_at	4.01771E-05	X	0	X	X	3	NR4A2
200047_s_at	4.02771E-05	X	0	X	X	3	0
200098_s_at	4.03771E-05	X	0	X	X	3	0
200625_s_at	4.04771E-05	X	0	X	X	3	HTR2A
204321_at	4.34506E-05	X	0	X	X	3	PPM1A
225864_at	5.61772E-05	X	0	X	X	3	ACAT1
211207_s_at	5.62062E-05	X	0	0	0	1	CDK10
221952_x_at	5.88854E-05	X	0	0	0	1	0
201481_s_at	5.89777E-05	X	0	0	0	1	HCLS1
210056_at	6.33692E-05	X	0	0	0	1	LOC100289577
1559496_at	6.34773E-05	X	0	0	X	2	OR4D2
1569200_at	6.35773E-05	X	0	0	X	2	EFNA5
1569477_at	6.36773E-05	X	0	0	X	2	KLC2
200089_s_at	6.37773E-05	X	0	0	X	2	PRKAB2
227276_at	6.40813E-05	X	0	0	0	1	PHYKPL
217622_at	6.48976E-05	X	X	0	0	2	ARHGEF7
225870_s_at	6.58773E-05	X	0	0	X	2	PFDN4
1562110_at	6.59773E-05	X	0	0	X	2	LINC00957
1563130_a_at	6.60773E-05	X	0	0	X	2	0
1563482_at	6.61773E-05	X	0	0	X	2	STK24
225719_s_at	6.81773E-05	X	0	0	X	2	0
229204_at	7.7558E-05	X	0	0	0	1	SPIN3
203853_s_at	8.02331E-05	X	0	0	0	1	FGF1
204718_at	8.91003E-05	X	0	X	X	3	LIN7B
214783_s_at	0.000100813	X	X	0	0	2	0
210513_s_at	0.00011901	X	0	0	0	1	PRO
1565614_at	0.000132341	X	0	0	0	1	0
233901_at	0.00014584	X	0	0	0	1	C10orf67
1558299_at	0.000149345	X	X	0	X	3	SST
201515_s_at	0.000161231	X	0	0	0	1	CDADC1
1554825_at	0.00018396	X	X	0	X	3	AHR
235975_at	0.00019901	X	0	0	0	1	LMO4
209316_s_at	0.000211492	X	X	X	X	4	0
241873_at	0.000234625	X	0	0	X	2	MYCBP2
201660_at	0.000236252	X	0	0	0	1	DT
204622_x_at	0.000248173	X	0	X	X	3	THSD7A
231049_at	0.000255924	X	0	0	0	1	SYTL2
239822_at	0.000267151	X	0	0	0	1	SYTL2
207135_at	0.000270877	X	0	0	0	1	ZNF230
1554769_at	0.000301643	X	X	0	X	3	CD3EAP
227728_at	0.000348418	X	0	0	0	1	LINC00868
205412_at	0.000411693	X	0	0	0	1	0
203468_at	0.00045157	X	0	X	X	3	SLIT2
210891_s_at	0.000470964	X	0	0	0	1	SRP72
202957_at	0.000490495	X	0	0	0	1	0
241904_at	0.000513422	X	0	0	0	1	0
1555377_at	0.000569835	X	0	0	0	1	DOK3
233814_at	0.00059652	X	0	0	0	1	0
218906_x_at	0.000601415	X	0	0	0	1	0
1556007_s_at	0.000657488	X	X	0	X	3	DENND5A

Sensitivity analysis result. (continued)

Probe ID	BF-value	CABk	EC	HIP	MTG-PC	DS	Symbol
1558027_s_at	0.000669004	X	0	0	0	1	CCDC155
232488_at	0.00068181	X	0	0	0	1	BLMH
239397_at	0.000754037	X	0	0	0	1	LH
205362_s_at	0.000879195	X	0	0	0	1	FYN
1564207_at	0.000982961	X	0	0	0	1	ATF6
243891_at	0.001003186	X	0	0	0	1	0
215188_at	0.001011148	X	0	0	0	1	MCM3AP
230333_at	0.001240065	X	0	0	0	1	ATP6V1G1
1555882_at	0.001276808	X	0	0	0	1	AMD1
205117_at	0.001317185	X	0	X	X	3	MAP1A
241957_x_at	0.001470117	X	0	0	0	1	MIF-AS1
1563881_at	0.001515133	X	X	X	X	4	CCDC117
226560_at	0.001662143	X	0	0	0	1	ZFAND6
229376_at	0.002324168	X	0	0	0	1	APC
223203_at	0.002418978	X	0	0	0	1	PCSK7
1553845_x_at	0.002542545	X	0	0	0	1	RAPGEF5
213921_at	0.002566735	X	0	X	0	2	APOC1
1555923_a_at	0.002730308	X	X	X	X	4	RPS11
233647_s_at	0.002734903	X	0	X	X	3	THTPA
202820_at	0.002807417	X	0	X	X	3	FAM178A
209204_at	0.002852804	X	0	0	0	1	ATP6V0E1
225923_at	0.003577085	X	X	X	X	4	MED25
1557623_at	0.003847887	X	0	0	0	1	0
201959_s_at	0.004155917	X	0	0	0	1	TFE3
235721_at	0.004199939	X	0	0	0	1	PRKCA
214920_at	0.004353403	X	0	0	0	1	LARS
225496_s_at	0.004580342	X	0	X	X	3	KLHL17
232914_s_at	0.004767688	X	0	0	0	1	SLC25A37
205791_x_at	0.004859877	X	0	0	0	1	CHD5
205264_at	0.005112205	X	0	0	0	1	ZC3H15
225880_at	0.005176354	X	0	0	X	2	RAB14
1556533_at	0.005251846	X	0	0	0	1	0
243314_at	0.005319749	X	0	0	0	1	HSP90AA1
209897_s_at	0.005441154	X	0	0	0	1	RPL27A
208095_s_at	0.00598977	X	0	0	0	1	CFL2
1555489_at	0.006361311	X	0	0	0	1	PTPRA
1558048_x_at	0.006723842	X	0	0	0	1	NR
200079_s_at	0.006949457	X	X	0	X	3	ITPRIPL2
223553_s_at	0.007121759	X	X	0	0	2	ZNF580
202759_s_at	0.00804011	X	0	0	0	1	RNASEH2C
1565852_at	0.008163459	X	0	0	0	1	HCFC1
1562201_x_at	0.008541514	X	X	0	X	3	0
212561_at	0.008826184	X	0	0	0	1	0
200685_at	0.009969508	X	X	X	X	4	NFIC
240950_s_at	0.01056431	X	0	0	0	1	SH3GLB2
202179_at	0.011825779	X	0	0	0	1	STO
206140_at	0.012278759	X	X	X	X	4	DMPK
210105_s_at	0.01244032	X	0	0	0	1	MRPL41
226941_at	0.012453317	X	0	0	0	1	SMARCA2
1562280_at	0.013148916	X	X	X	X	4	0
1558345_a_at	0.014496547	X	X	X	X	4	ZNF213
224375_at	0.015557395	X	0	X	0	2	BUB3
212269_s_at	0.016473826	X	0	0	0	1	CEP170B
208737_at	0.018450689	X	0	0	0	1	CEP97
201197_at	0.018896935	X	0	0	0	1	GNB1L

Sensitivity analysis result. (continued)

Probe ID	BF-value	CABk	EC	HIP	MTG-PC	DS	Symbol
215391_at	0.025927655	X	0	0	0	1	CRLS1
1556316_s_at	0.027708828	X	0	0	0	1	0
225644_at	0.035880774	X	0	0	0	1	VPS13C
221613_s_at	0.04159524	X	0	0	0	1	ZNF561
203525_s_at	0.045024916	X	0	0	0	1	NUCKS1
232248_at	0.046657157	X	0	0	0	1	N4BP1
225750_at	0.059320315	X	X	0	X	3	0
204680_s_at	0.06096767	X	X	X	X	4	ZBED6
213553_x_at	0.062918755	X	0	0	0	1	SPG7
213350_at	0.072927203	X	0	0	0	1	0
218540_at	0.073260507	X	0	0	0	1	DNAJC18
203482_at	0.07333948	X	0	0	X	2	0
214149_s_at	0.089692728	X	0	0	0	1	CWF19L2
1553993_s_at	0.098092936	X	0	0	0	1	0
207909_x_at	0.10784315	X	0	0	0	1	PCBP1-AS1
212457_at	0.108496219	X	0	0	0	1	0
213093_at	0.112511026	X	0	X	X	3	EGFR
222428_s_at	0.119814104	X	0	X	X	3	C16orf52
229792_at	0.121872102	X	0	0	0	1	LOC642361
226179_at	0.130745814	X	0	0	0	1	0
217250_s_at	0.136631051	X	0	X	X	3	ZNF827
201593_s_at	0.142087053	X	X	0	0	2	0
211503_s_at	0.147339432	X	0	0	0	1	DVL2
211996_s_at	0.156751107	X	X	0	0	2	EPC1
211968_s_at	0.165816317	X	0	0	0	1	0
203034_s_at	0.186475028	X	0	0	0	1	GAS2L1
224352_s_at	0.191145471	X	0	0	0	1	0
213795_s_at	0.206711343	X	0	0	0	1	0
209982_s_at	0.208576506	X	0	0	0	1	0
228074_at	0.256395169	X	0	0	0	1	0
220748_s_at	0.320551736	X	0	0	0	1	LOC646214
226453_at	0.320891439	X	0	0	0	1	0
202473_x_at	0.38741949	X	0	0	0	1	0
231034_s_at	0.551911998	X	0	0	0	1	DTNA
237674_at	0.555721628	X	0	0	0	1	WWC1
213298_at	0.602952484	X	0	0	X	2	0
224432_at	0.606434894	X	0	0	0	1	HIVEP3
231969_at	0.691670928	X	0	0	0	1	PIK3C2A
37996_s_at	0.704281792	X	X	X	X	4	RIMS1
225423_x_at	0.705780972	X	0	0	0	1	0
217707_x_at	0.71787385	X	X	0	0	2	0
223929_s_at	0.860777115	X	0	0	0	1	NDUFAF7
227207_x_at	0.910056486	X	0	0	0	1	0
209974_s_at	0.923205984	X	0	0	0	1	YTHDC1
213242_x_at	0.995833381	X	0	0	0	1	0
235918_x_at	1.771538992	X	0	0	X	2	MRPS5
220762_s_at	1.771538992	X	0	0	0	1	ELK4
241741_at	1.771538992	X	X	X	X	4	MZF1
242343_x_at	1.771538992	X	X	X	X	4	ZNF721
232386_at	1.771538992	X	0	X	X	3	0
235200_at	1.771538992	X	X	X	0	3	ING5
229353_s_at	1.771538992	X	X	0	0	2	LMF2
32069_at	1.771538992	X	X	X	X	4	0
238812_at	1.771538992	X	X	X	X	4	AKAP8L
243648_at	1.771538992	X	0	X	X	3	CFLAR

Sensitivity analysis result. (continued)

Probe ID	BF-value	CABk	EC	HIP	MTG-PC	DS	Symbol
230885_at	1.771538992	X	X	X	X	4	FB
237018_at	1.771538992	X	X	0	X	3	RPL7A
238115_at	1.771538992	X	0	X	X	3	GOLGA2
229264_at	1.771538992	X	0	0	X	2	TRIM46
237040_at	1.771538992	X	0	X	X	3	0
237768_x_at	1.771538992	X	X	X	X	4	0
235482_at	1.771538992	X	0	X	0	2	TRIM9
239497_at	1.771538992	X	X	0	X	3	0
1565483_at	1.771538992	X	0	0	0	1	0
230296_at	1.781538992	X	X	0	X	3	EPM2AIP1
228839_s_at	1.791538992	X	X	X	X	4	0
243329_at	1.801538992	X	X	0	X	3	0
243617_at	1.811538992	X	0	X	X	3	0
239096_at	1.821538992	X	X	X	0	3	0
57532_at	1.831538992	X	X	X	0	3	0
234969_s_at	1.841538992	X	X	X	X	4	0
243442_x_at	1.851538992	X	X	0	0	2	WDR1
31874_at	1.861538992	X	X	0	0	2	0
236752_at	1.871538992	X	0	X	X	3	CIRH1A
244778_x_at	1.881538992	X	0	X	X	3	ANAPC16
244803_at	1.891538992	X	X	X	0	3	0
237937_x_at	1.901538992	X	X	0	0	2	0
236283_x_at	1.911538992	X	X	X	X	4	0
236488_s_at	1.921538992	X	X	X	X	4	EP400
228858_at	1.931538992	X	X	0	0	2	CAPZB
227084_at	1.941538992	X	0	X	X	3	SMAD1
236725_at	1.951538992	X	X	0	0	2	0
232288_at	1.961538992	X	X	0	X	3	PRR11
235122_at	1.971538992	X	X	0	X	3	0
241905_at	1.981538992	X	0	0	0	1	LOC286437
231986_at	1.991538992	X	0	X	X	3	LOC100289494
229315_at	2.001538992	X	0	X	X	3	FAM19A1
239102_s_at	2.006538112	X	X	0	X	3	NADK2
230379_x_at	2.007538112	X	X	0	X	3	0
243827_at	2.011538992	X	X	X	X	4	0
228556_at	2.021538992	X	0	X	X	3	0
242712_x_at	2.028537672	X	X	X	X	4	ADRB1
237560_at	2.031538992	X	X	X	X	4	ACACB
238761_at	2.041538992	X	0	X	X	3	0
40569_at	2.049537232	X	0	X	X	3	CEP85L
228029_at	2.051538992	X	0	0	X	2	KIAA2018
239469_at	2.061538992	X	0	0	X	2	0
228287_at	2.070536792	X	X	X	X	4	0
31837_at	2.071538992	X	X	0	0	2	ABHD2
242240_at	2.082538992	X	X	X	0	3	0
240554_at	2.091536352	X	0	X	X	3	CMBL
239629_at	2.093538992	X	X	X	X	4	LOC222070
242829_x_at	2.104538992	X	X	X	X	4	RUFY3
234873_x_at	2.112535912	X	0	X	0	2	0
35436_at	2.115538992	X	0	X	0	2	0
238147_at	2.126538992	X	X	0	X	3	KIAA0368
235964_x_at	2.133535472	X	0	X	X	3	0
239678_at	2.137538992	X	X	X	X	4	DIS3L2
230280_at	2.148538992	X	0	X	X	3	0
233037_at	2.154535032	X	0	X	X	3	0

Sensitivity analysis result. (continued)

Probe ID	BF-value	CABk	EC	HIP	MTG-PC	DS	Symbol
230790_x_at	2.159538992	X	X	X	X	4	CNOT3
227847_at	2.170538992	X	0	X	X	3	NUMBL
240758_at	2.175534592	X	0	X	X	3	0
244859_at	2.181538992	X	0	0	X	2	DHFRL1
239179_at	2.192538992	X	0	X	X	3	CLEC2L
228487_s_at	2.196534152	X	0	X	X	3	0
231281_at	2.203538992	X	X	X	X	4	0
236153_at	2.214538992	X	0	0	0	1	0
240282_at	2.217533712	X	0	0	X	2	0
233405_at	2.225538992	X	X	0	X	3	ACACB
230656_s_at	2.236538992	X	0	X	X	3	0
229145_at	2.238533272	X	X	X	X	4	FDFT1
243593_s_at	2.247538992	X	X	X	0	3	SYS1
244511_at	2.258538992	X	0	X	X	3	SGSM2
233702_x_at	2.259532832	X	0	X	X	3	CAPN1
230629_s_at	2.269538992	X	X	0	0	2	REV3L
37012_at	2.280532392	X	0	X	0	2	SLC22A3
227798_at	2.280538992	X	0	X	X	3	0
243025_at	2.291538992	X	X	X	X	4	PPP2R2C
232215_x_at	2.301531952	X	0	X	X	3	C3orf70
230713_at	2.302538992	X	X	X	X	4	B2M
239762_at	2.313538992	X	X	0	0	2	KDM4B
231387_at	2.322531512	X	X	X	X	4	IAH1
230923_at	2.324538992	X	0	X	X	3	DCAF8
228594_at	2.335538992	X	X	X	X	4	0
234723_x_at	2.343531072	X	0	X	X	3	0
241391_at	2.346538992	X	X	X	X	4	0
235730_at	2.357538992	X	0	X	X	3	0
229309_at	2.364530632	X	0	X	X	3	ACSL6
43427_at	2.368538992	X	0	X	X	3	TGS1
227772_at	2.379538992	X	0	X	X	3	TTC27
228007_at	2.385530192	X	0	X	X	3	NUDT7
227435_at	2.390538992	X	X	X	X	4	C11orf1
231329_at	2.401538992	X	X	X	X	4	0
244027_at	2.406529752	X	0	X	X	3	EIF3B
63825_at	2.412538992	X	0	X	X	3	0
240018_at	2.423538992	X	0	X	0	2	ZNF83
234981_x_at	2.427529312	X	0	X	X	3	0
51774_s_at	2.434538992	X	X	0	0	2	CHRNB2
227802_at	2.445538992	X	0	X	X	3	SP4
237798_at	2.448528872	X	X	0	X	3	SEC22C
241786_at	2.456538992	X	X	0	X	3	CHRM1
236368_at	2.467538992	X	0	X	X	3	MGC70870
241893_at	2.469528432	X	X	X	0	3	EPHA4
238602_at	2.478538992	X	X	X	X	4	KCNIP4
236934_at	2.489538992	X	0	X	X	3	SEMA3F
236139_at	2.490527992	X	0	X	X	3	DENR
229143_at	2.500538992	X	X	X	X	4	ZFAND6
242195_x_at	2.511527552	X	X	X	X	4	ANKDD1A
237107_at	2.511538992	X	0	0	X	2	CACNA1C
241727_x_at	2.522538992	X	X	X	X	4	0
232173_at	2.532527112	X	0	X	X	3	ITPKB
228974_at	2.533538992	X	0	X	X	3	0
244045_at	2.544538992	X	0	X	0	2	0
235887_at	2.553526672	X	0	X	X	3	FAM85A

Sensitivity analysis result. (continued)

Probe ID	BF-value	CABk	EC	HIP	MTG-PC	DS	Symbol
237600_at	2.555538992	X	X	X	X	4	0
49452_at	2.566538992	X	0	X	X	3	OTUB1
243525_at	2.574526232	X	0	X	X	3	0
241954_at	2.577538992	X	X	X	X	4	MED16
238470_at	2.588538992	X	0	X	X	3	ITPRIPL2
36129_at	2.595525792	X	X	X	X	4	0
232012_at	2.599538992	X	X	0	X	3	SLC32A1
238736_at	2.610538992	X	0	0	X	2	0
242578_x_at	2.616525352	X	0	X	X	3	0
238108_at	2.621538992	X	0	X	0	2	NME7
228010_at	2.632538992	X	0	0	X	2	ATAD1
235562_at	2.637524912	X	X	X	X	4	ZFR
232311_at	2.643538992	X	X	X	X	4	UBE3A
235789_at	2.654538992	X	0	X	X	3	KMT2D
230621_at	2.658524472	X	0	X	X	3	0
243318_at	2.665538992	X	0	0	X	2	ANKRD34A
244610_x_at	2.676538992	X	0	X	X	3	0
233319_x_at	2.679524032	X	X	X	X	4	SRSF4
236949_at	2.687538992	X	X	X	X	4	0
227663_at	2.698538992	X	X	X	X	4	0
229725_at	2.700523592	X	0	X	X	3	NFATC2IP
238346_s_at	2.709538992	X	0	X	X	3	0
244248_at	2.720538992	X	X	X	0	3	FNIP1
228855_at	2.721523152	X	0	X	X	3	ANK2
231530_s_at	2.731538992	X	0	X	X	3	MALAT1
243275_at	2.742522712	X	X	X	X	4	0
236274_at	2.742538992	X	0	X	X	3	KHSRP
235288_at	2.753538992	X	X	X	X	4	GLG1
236429_at	2.763522272	X	0	X	X	3	0
232797_at	2.764538992	X	0	X	X	3	RAI1
241389_at	2.775538992	X	0	X	X	3	LINC00657
236265_at	2.784521832	X	0	0	X	2	PITRM1
236268_at	2.786538992	X	0	X	X	3	KTN1
231783_at	2.797538992	X	X	0	X	3	YLPM1
242136_x_at	2.805521392	X	0	X	X	3	SFSWAP
229374_at	2.808538992	X	0	X	X	3	NAV2
236783_at	2.819538992	X	0	X	X	3	RTN3
35666_at	2.826520952	X	X	X	X	4	PEBP1
231896_s_at	2.830538992	X	X	0	X	3	0
239757_at	2.841538992	X	X	X	0	3	EAPP
229497_at	2.847520512	X	X	X	X	4	0
242973_at	2.852538992	X	X	X	X	4	NAV1
234989_at	2.863538992	X	X	X	X	4	UBE2W
235213_at	2.868520072	X	0	X	0	2	JAK1
240005_at	2.874538992	X	0	X	X	3	PNISR
244673_at	2.885538992	X	0	X	X	3	0
238716_at	2.889519632	X	0	X	X	3	0
242239_at	2.896538992	X	X	X	X	4	NDRG4
38710_at	2.907538992	X	0	X	X	3	LZTS2
243459_x_at	2.910519192	X	0	X	0	2	ELAVL3
43544_at	2.918538992	X	X	X	X	4	PEBP1
227792_at	2.929538992	X	0	X	X	3	ZNF827
234578_at	2.931518752	X	0	X	X	3	NSL1
240532_at	2.940538992	X	0	X	X	3	0
242688_at	2.951538992	X	X	X	X	4	0

Sensitivity analysis result. (continued)

Probe ID	BF-value	CABk	EC	HIP	MTG-PC	DS	Symbol
234081_at	2.952518312	X	0	X	X	3	MTMR4
227556_at	2.962538992	X	0	X	X	3	RB1
227585_at	2.973517872	X	X	X	0	3	0
33148_at	2.973538772	X	0	X	X	3	RBM39
234163_at	2.984538552	X	X	X	X	4	SLCO3A1
227527_at	2.994517432	X	0	X	X	3	0
243561_at	2.995538332	X	X	X	X	4	BLCAP
232735_at	3.015516992	X	0	X	X	3	C16orf62
242407_at	3.036516552	X	0	X	X	3	MADD
242837_at	3.057516112	X	0	X	X	3	CSRNP1
240331_at	3.078515672	X	0	X	X	3	NDUFA6
233908_x_at	3.099515232	X	0	X	X	3	0
238130_at	3.120514792	X	X	X	X	4	PTAR1
225961_at	3.499099974	X	X	X	X	4	FBRSL1
226086_at	3.562772246	X	0	X	X	3	PPP1R11
226101_at	3.600056377	X	0	X	X	3	SREBF2
226112_at	5.64772E-05	0	0	X	X	2	0
226143_at	1.93842E-06	0	X	0	X	2	SEC62
226169_at	5.25703E-05	0	X	X	X	3	0
226195_at	0.006560981	0	X	X	X	3	DNASE1
226223_at	5.65772E-05	0	0	X	X	2	0
226484_at	0.000134175	0	X	0	X	2	0
226501_at	5.67772E-05	0	0	X	X	2	MKNK1
226544_x_at	5.68772E-05	0	0	X	X	2	RBM27
226573_at	6.21773E-05	0	0	X	X	2	0
226625_at	5.69772E-05	0	0	X	X	2	RAB18
226690_at	5.70772E-05	0	0	X	X	2	0
226718_at	5.71772E-05	0	0	X	X	2	0
226791_at	6.22773E-05	0	0	X	X	2	0
226886_at	0.005851786	0	X	X	X	3	0
226892_at	0.014555841	0	0	X	X	2	PTPN13
226898_s_at	5.73772E-05	0	0	X	X	2	0
1552302_at	0.000622706	0	X	0	0	1	ATR
1552735_at	5.78772E-05	0	0	X	0	1	0
1553107_s_at	0.000142302	0	X	0	0	1	0
1553172_at	0.001707674	0	X	0	0	1	0
1553681_a_at	0.000355451	0	X	0	0	1	0
1553909_x_at	7.99312E-05	0	X	0	0	1	DAAM1
1553960_at	5.79772E-05	0	0	X	0	1	SMURF2
1553984_s_at	0.01465517	0	X	0	0	1	GLIS1
1554116_s_at	0.000947325	0	X	0	0	1	0
1554565_x_at	0.000273438	0	X	0	0	1	0
1554770_x_at	0.000334463	0	X	0	0	1	0
1554963_at	0.009028291	0	X	0	0	1	0
1555363_s_at	0.005548915	0	X	0	0	1	SAFB2
1555579_s_at	5.80772E-05	0	0	X	0	1	PPARD
1555881_s_at	1.15901E-05	0	X	0	0	1	MADD
1555894_s_at	0.000214914	0	X	0	0	1	WAS
1556069_s_at	0.009438387	0	X	0	0	1	SDCCAG3
1556336_at	0.010961736	0	X	0	0	1	GAK
1556416_s_at	0.003466097	0	X	0	0	1	KDM6B
1556641_at	0.000355493	0	X	0	0	1	TJAP1
1556682_s_at	0.021545109	0	X	0	0	1	WIZ
1556950_s_at	0.007056233	0	X	0	0	1	COL5A3
1557012_a_at	0.017161249	0	X	0	0	1	AMBRA1

Sensitivity analysis result. (continued)

Probe ID	BF-value	CABk	EC	HIP	MTG-PC	DS	Symbol
1557384_at	0.053487247	0	X	0	0	1	ARHGEF40
1557507_at	0.000684526	0	X	0	0	1	LOC222070
1557718_at	0.003060578	0	X	0	0	1	ENGASE
1557737_s_at	0.044086983	0	X	0	0	1	KCNE4
1557780_at	0.036855511	0	X	0	0	1	RASEF
1557814_a_at	0.013272055	0	X	0	0	1	0
1557996_at	0.023020069	0	X	0	0	1	0
1558078_at	0.002042186	0	X	0	0	1	FLJ35934
1558426_x_at	0.00046153	0	X	0	0	1	KDSR
1558592_at	0.00067536	0	X	0	0	1	0
1558692_at	0.047622726	0	X	X	0	2	0
1558740_s_at	0.002015811	0	X	0	0	1	0
1558783_at	0.006316994	0	X	0	0	1	NEDD1
1559023_a_at	0.008734763	0	X	0	0	1	AKT2
1559060_a_at	0.00270604	0	X	X	0	2	SNRPN
1559259_at	0.002508714	0	X	X	0	2	0
1559402_a_at	0.002791753	0	X	0	0	1	LOC285949
1559479_at	0.001308217	0	X	0	0	1	AGAP4
1559618_at	0.000539918	0	X	0	0	1	PPP1R11
1560199_x_at	0.00135891	0	X	0	0	1	CADPS
1560445_x_at	0.001841521	0	X	0	0	1	GABRG2
1560775_at	0.013114754	0	X	0	0	1	0
1560982_at	0.001900874	0	X	X	0	2	PPP6R2
1561167_at	0.021513573	0	X	0	0	1	ANAPC5
1561362_at	0.000294668	0	X	0	0	1	CAP1
1561642_at	0.005975484	0	X	0	0	1	YWHAQ
1562013_a_at	0.006341259	0	X	0	0	1	HSPA1A
1562144_at	0.000280563	0	X	0	0	1	BCLAF1
1562529_s_at	0.086032289	0	X	0	0	1	TRIM14
1564378_a_at	4.63638E-05	0	X	0	0	1	C3orf38
1566079_at	0.002047843	0	X	0	0	1	CHD4
1566303_s_at	3.23274E-05	0	0	X	0	1	SEC11A
1566848_x_at	0.000882539	0	X	0	0	1	PLEKHB2
1566887_x_at	0.001402054	0	X	X	0	2	FAM127A
1568916_at	0.00022314	0	X	X	0	2	SUPT7L
1569069_s_at	0.027491457	0	X	0	0	1	KIF5B
1569126_at	0.001871423	0	X	0	0	1	ACAA1
1569129_s_at	4.0809E-05	0	X	0	0	1	IDH3A
1569409_x_at	0.002597212	0	X	0	0	1	AP2S1
1569597_at	0.015598105	0	X	0	0	1	MAML1
1569661_at	0.001193615	0	X	0	0	1	PFKFB3
1570511_at	0.004928538	0	X	0	0	1	IDH3G
200001_at	0.010575615	0	X	0	0	1	SNRPB2
200006_at	9.58176E-05	0	X	0	0	1	LEPROTL1
200009_at	0.000119411	0	X	0	0	1	TOB1
200064_at	0.000105884	0	X	0	0	1	UROS
200072_s_at	0.008245468	0	X	X	0	2	BMS1
200082_s_at	0.010108959	0	X	0	0	1	EEF1D
200099_s_at	5.81772E-05	0	0	X	0	1	TRAPPC12
200623_s_at	5.82772E-05	0	0	X	0	1	RB1
200629_at	0.000678094	0	X	0	0	1	BCL6
200631_s_at	0.004889418	0	X	0	0	1	FB
200642_at	0.000178484	0	X	0	0	1	ARHGAP32
200647_x_at	0.000193384	0	X	0	0	1	LPL
200650_s_at	0.00159714	0	X	0	0	1	TCF4

Sensitivity analysis result. (continued)

Probe ID	BF-value	CABk	EC	HIP	MTG-PC	DS	Symbol
200657_at	0.000186538	0	X	0	0	1	NQO2
200658_s_at	0.000580105	0	X	0	0	1	KIF1A
200693_at	4.05771E-05	0	0	X	X	2	RAP1GAP
200695_at	0.000105889	0	X	0	0	1	BTN2A1
200703_at	0.001426002	0	X	0	0	1	SLC17A7
200705_s_at	5.83772E-05	0	0	X	0	1	CHKA
200735_x_at	0.00213711	0	X	0	0	1	BRPF1
200750_s_at	0.000584034	0	X	0	0	1	PPIC
200754_x_at	7.28772E-05	0	X	X	X	3	NAP1L1
200775_s_at	0.003852231	0	X	0	X	2	LSM7
200786_at	0.001799097	0	X	0	0	1	L1CAM
200792_at	5.67176E-05	0	X	0	0	1	DNAJC6
200799_at	4.06771E-05	0	0	X	0	1	ORC5
200812_at	0.004172767	0	X	0	0	1	RPH3A
200813_s_at	0.00143306	0	X	X	0	2	AMPH
200818_at	8.57126E-05	0	X	0	0	1	PIP4K2A
200831_s_at	0.003043478	0	X	0	0	1	ANGPT1
200862_at	0.000718481	0	X	0	0	1	HIST1H4C
200876_s_at	0.002005306	0	X	0	0	1	SO
200883_at	0.001161378	0	X	X	0	2	SLMO1
200901_s_at	8.27193E-05	0	X	0	0	1	PPEF1
200903_s_at	0.006467137	0	X	0	0	1	0
200914_x_at	3.73591E-06	0	X	0	0	1	SPTAN1
200932_s_at	0.000160117	0	X	0	X	2	AP3D1
200955_at	0.000120376	0	X	0	0	1	RBM39
200960_x_at	0.000121944	0	X	0	0	1	DUT
200982_s_at	0.002436438	0	X	0	0	1	ATP5G1
201000_at	0.003099536	0	X	0	0	1	POLR2C
201024_x_at	0.000361042	0	X	0	0	1	0
201030_x_at	0.000158597	0	X	0	0	1	UQCRB
201032_at	2.26583E-05	0	X	0	0	1	TUBA1C
201072_s_at	5.51506E-05	0	X	0	0	1	VPS45
201076_at	0.002325294	0	X	0	0	1	NFIB
201083_s_at	4.07771E-05	0	0	X	X	2	EFHD1
201090_x_at	0.000154268	0	X	X	0	2	GRB10
201093_x_at	0.008864222	0	X	0	0	1	0
201119_s_at	0.027114157	0	X	0	0	1	PRUNE
201123_s_at	0.004651099	0	X	X	0	2	MRPL9
201145_at	0.009303074	0	X	X	0	2	S100B
201160_s_at	6.38773E-05	0	0	0	X	1	CB
201174_s_at	0.00012536	0	X	0	0	1	FAM220A
201182_s_at	4.08771E-05	0	0	X	X	2	ZNF330
201191_at	0.000972103	0	X	0	0	1	GABBR2
201199_s_at	0.00086347	0	X	0	0	1	ATP1A1
201216_at	0.006221225	0	X	X	0	2	LOC100287552
201217_x_at	0.002415825	0	X	X	X	3	0
201241_at	0.000557182	0	X	0	0	1	IQGAP1
201267_s_at	0.000992651	0	X	0	0	1	AP2A2
201272_at	5.84772E-05	0	0	X	0	1	COL4A2
201274_at	0.002164824	0	X	0	0	1	PNN
201290_at	4.09771E-05	0	0	X	X	2	PMPCA
201313_at	0.000784385	0	X	X	0	2	MTUS1
201322_at	0.001988759	0	X	0	0	1	RBFO
201330_at	0.0112829	0	X	0	0	1	MED13L
201341_at	0.000281717	0	X	0	0	1	COQ9

Sensitivity analysis result. (continued)

Probe ID	BF-value	CABk	EC	HIP	MTG-PC	DS	Symbol
201387_s_at	0.001023108	0	X	0	0	1	MYH10
201390_s_at	0.00110521	0	X	0	0	1	GRPEL1
201410_at	4.10771E-05	0	0	X	X	2	PI4KA
201441_at	0.001315733	0	X	0	X	2	RPL17
201475_x_at	0.000346201	0	X	0	0	1	SN
201493_s_at	5.85772E-05	0	0	X	0	1	TROVE2
201499_s_at	0.016070586	0	X	0	0	1	0
201502_s_at	5.86772E-05	0	0	X	0	1	NFIB
201503_at	0.0082839	0	X	0	0	1	LOC100272216
201509_at	0.079403357	0	X	X	0	2	0
201520_s_at	0.008105895	0	X	0	0	1	LYRM9
201522_x_at	0.000439896	0	X	0	0	1	MDH2
201524_x_at	0.005801016	0	X	0	0	1	UBE2I
201548_s_at	0.008520804	0	X	0	X	2	SN
201565_s_at	0.026386739	0	X	X	X	3	SO
201586_s_at	6.39773E-05	0	0	0	X	1	ATRNL1
201622_at	0.001102468	0	X	0	X	2	PTPRD
201628_s_at	0.002773808	0	X	0	0	1	EEF1D
201632_at	0.003344118	0	X	0	0	1	FB
201648_at	8.73029E-06	0	X	X	0	2	CU
201652_at	0.001616678	0	X	0	0	1	LOC100170939
201672_s_at	0.011367004	0	X	0	0	1	ATP9B
201709_s_at	0.000929848	0	X	X	X	3	0
201725_at	0.000270719	0	X	0	0	1	0
201740_at	0.002857108	0	X	0	0	1	0
201754_at	0.022957546	0	X	0	0	1	FB
201757_at	0.001041115	0	X	0	0	1	0
201804_x_at	0.011530663	0	X	0	0	1	0
201810_s_at	0.004429599	0	X	0	X	2	TRIM44
201825_s_at	0.007639684	0	X	0	0	1	BACE1
201828_x_at	4.11771E-05	0	0	X	X	2	AFTPH
201836_s_at	4.12771E-05	0	0	X	X	2	SAE1
201856_s_at	0.000835899	0	X	0	0	1	VPS51
201880_at	0.018478256	0	X	0	0	1	C14orf159
201892_s_at	0.002677939	0	X	0	0	1	PSENEN
201922_at	0.025165321	0	X	0	0	1	FNDC4
201947_s_at	0.000113805	0	X	0	0	1	0
201964_at	0.001883641	0	X	0	0	1	OPN3
201991_s_at	4.13771E-05	0	0	X	X	2	EMC9
202000_at	2.64779E-05	0	X	0	0	1	SBF2
202025_x_at	4.14771E-05	0	0	X	X	2	FAM173A
202066_at	4.15771E-05	0	0	X	X	2	MMADHC
202070_s_at	4.16771E-05	0	0	X	X	2	HAUS2
202077_at	0.005365909	0	X	0	0	1	CRYBA2
202120_x_at	4.17771E-05	0	0	X	X	2	ZBB
202121_s_at	0.001399366	0	X	0	0	1	DEPDC1
202154_x_at	0.00258081	0	X	0	0	1	FAR2
202178_at	0.001537846	0	X	X	0	2	ENSA
202184_s_at	5.87772E-05	0	0	X	0	1	CUTA
202201_at	6.40773E-05	0	0	0	X	1	PARD3
202230_s_at	0.000189525	0	X	0	0	1	PPP2R2D
202242_at	0.000234773	0	X	X	0	2	ORAI3
202243_s_at	0.002911991	0	X	0	0	1	GUCY1A3
202244_at	0.010399997	0	X	0	0	1	0
202268_s_at	0.001316385	0	X	0	0	1	0

Sensitivity analysis result. (continued)

Probe ID	BF-value	CABk	EC	HIP	MTG-PC	DS	Symbol
202298_at	0.002293178	0	X	0	X	2	LIMA1
202307_s_at	6.62773E-05	0	0	0	X	1	KATNBL1
202309_at	6.63773E-05	0	0	0	X	1	BHLHE41
202343_x_at	0.000468755	0	X	0	0	1	RF
202353_s_at	7.99303E-05	0	X	0	0	1	AK3
202360_at	4.18771E-05	0	0	X	X	2	MALAT1
202363_at	0.000534256	0	X	X	0	2	DNAJC5
202416_at	0.000148744	0	X	X	X	3	ERLEC1
202464_s_at	4.19771E-05	0	0	X	X	2	CCAR1
202467_s_at	0.000161177	0	X	0	0	1	GNAQ
202471_s_at	4.20771E-05	0	0	X	X	2	SMARCC1
202505_at	4.21771E-05	0	0	X	X	2	TMEM205
202507_s_at	0.001881368	0	X	0	0	1	LRRC37A16P
202517_at	0.000200462	0	X	0	0	1	RABEP1
202552_s_at	6.64773E-05	0	0	0	X	1	SMAD5
202573_at	0.001233803	0	X	0	0	1	PNKD
202594_at	4.22771E-05	0	0	X	X	2	GPR89A
202623_at	7.12498E-06	0	X	0	0	1	CEP41
202650_s_at	0.006805594	0	X	0	0	1	SLC7A2
202660_at	6.41773E-05	0	0	0	X	1	STAT2
202675_at	0.002905341	0	X	0	0	1	UHRF1
202683_s_at	9.03445E-05	0	X	0	0	1	FAM84B
202698_x_at	6.43674E-05	0	X	X	0	2	PRKCE
202704_at	4.23771E-05	0	0	X	X	2	SGCB
202717_s_at	0.000538643	0	X	X	X	3	0
202724_s_at	5.88773E-05	0	0	X	0	1	0
202736_s_at	0.086226116	0	X	0	0	1	ASH1L
202773_s_at	5.22461E-06	0	X	0	X	2	0
202795_x_at	0.020147898	0	X	0	0	1	BLOC1S5
202858_at	0.000462229	0	X	X	X	3	TGFBR3
202868_s_at	0.000573659	0	X	X	X	3	AMIGO1
202897_at	0.000387156	0	X	0	0	1	C10orf12
202916_s_at	0.002834025	0	X	0	0	1	0
202920_at	4.12766E-07	0	X	0	0	1	SFPQ
202926_at	0.001246898	0	X	0	0	1	MOB3C
202941_at	3.83866E-05	0	X	0	0	1	GABRD
202967_at	0.000187435	0	X	0	X	2	TET2
202974_at	0.013450621	0	X	X	X	3	CLCN7
203000_at	0.002861204	0	X	0	0	1	PCDHGA4
203029_s_at	0.070196916	0	X	0	0	1	SN
203031_s_at	4.24771E-05	0	0	X	X	2	PTPRM
203033_x_at	0.001153232	0	X	X	0	2	RPS3A
203082_at	4.25771E-05	0	0	X	X	2	ZNF652
203113_s_at	4.26771E-05	0	0	X	X	2	CALM3
203122_at	4.27771E-05	0	0	X	X	2	EEF1B2
203132_at	2.00676E-05	0	0	X	X	2	AKR1B1
203140_at	4.29771E-05	0	0	X	X	2	PUM2
203147_s_at	4.0791E-05	0	X	0	0	1	NFKBIA
203157_s_at	0.000535782	0	X	0	0	1	FO
203160_s_at	0.000714804	0	X	0	0	1	RREB1
203173_s_at	2.44181E-05	0	0	0	X	1	RGS4
203227_s_at	0.001617991	0	X	0	0	1	VCAN
203230_at	0.000336146	0	X	0	0	1	PCSK2
203255_at	4.30771E-05	0	0	X	0	1	TFRC
203261_at	0.000826409	0	X	0	0	1	GHITM

Sensitivity analysis result. (continued)

Probe ID	BF-value	CABk	EC	HIP	MTG-PC	DS	Symbol
203266_s_at	0.001061502	0	X	0	0	1	ASMTL
203324_s_at	0.000468963	0	X	0	X	2	CCDC28A
203340_s_at	0.000129147	0	X	0	0	1	PNISR
203413_at	0.001028453	0	X	0	0	1	NOTCH2
203431_s_at	4.31771E-05	0	0	X	X	2	ARID5B
203442_x_at	0.011857879	0	X	0	X	2	P
203484_at	0.023715836	0	X	0	0	1	CTSB
203518_at	0.000386526	0	X	0	X	2	ZNF423
203549_s_at	4.32771E-05	0	0	X	X	2	TRAPPC10
203603_s_at	6.42773E-05	0	0	0	X	1	MAP2K3
203604_at	0.000138684	0	X	0	0	1	DD
203605_at	0.003272112	0	X	0	0	1	F
203704_s_at	5.89773E-05	0	0	X	0	1	PRC1
203721_s_at	0.009818264	0	X	0	0	1	C12orf10
203738_at	0.000712145	0	X	0	0	1	NUP43
203752_s_at	0.003519209	0	X	X	0	2	AGFG2
203753_at	4.33771E-05	0	0	X	X	2	BACE2
203762_s_at	0.010129245	0	X	0	0	1	MRPL37
203773_x_at	0.005899031	0	X	X	X	3	PACSIN1
203798_s_at	0.000356424	0	X	X	0	2	DRAM2
203814_s_at	4.34771E-05	0	0	X	X	2	RBMS2
203816_at	0.002287259	0	X	0	0	1	MANEAL
203839_s_at	0.005447321	0	X	0	0	1	CTTN
203849_s_at	4.35771E-05	0	0	X	X	2	USP42
203861_s_at	0.002101362	0	X	0	0	1	LCOR
203898_at	6.66773E-05	0	0	0	X	1	RBM27
203907_s_at	0.00414854	0	X	0	0	1	DIRAS1
203911_at	4.36771E-05	0	0	X	X	2	KIFC2
203944_x_at	4.37771E-05	0	0	X	0	1	ANKS1B
203956_at	8.99214E-05	0	X	X	X	3	SO
204050_s_at	0.00499733	0	X	0	0	1	PRKG1
204075_s_at	0.00452735	0	X	0	0	1	SLC25A29
204081_at	0.011525958	0	X	0	0	1	0
204156_at	0.002320474	0	X	0	0	1	B3GAT2
204206_at	0.002322359	0	X	0	0	1	GALNT15
204229_at	4.38771E-05	0	0	X	X	2	CIT
204260_at	0.002317452	0	X	0	0	1	ZNF827
204266_s_at	4.39771E-05	0	0	X	X	2	CC2D1A
204338_s_at	5.90773E-05	0	0	X	0	1	PPA2
204365_s_at	0.005613386	0	X	0	0	1	0
204372_s_at	1.34845E-06	0	X	0	0	1	0
204449_at	0.004357496	0	X	0	0	1	RPL4
204465_s_at	0.001135759	0	X	0	0	1	YB
204481_at	4.40771E-05	0	0	X	X	2	SFPQ
204517_at	4.41771E-05	0	0	X	X	2	BLVRB
204528_s_at	4.42771E-05	0	0	X	X	2	ITPR2
204546_at	0.000711072	0	X	0	0	1	ZEB2
204559_s_at	4.43771E-05	0	0	X	X	2	CO
204584_at	4.44771E-05	0	0	X	X	2	TGFBR3
204587_at	0.018455552	0	X	X	0	2	AAK1
204619_s_at	5.91773E-05	0	0	X	0	1	0
204663_at	0.002296899	0	X	0	X	2	0
204685_s_at	0.009475286	0	X	X	0	2	CFDP1
204720_s_at	4.45771E-05	0	0	X	X	2	SPAG9
204729_s_at	0.018757615	0	X	0	0	1	GRHPR

Sensitivity analysis result. (continued)

Probe ID	BF-value	CABk	EC	HIP	MTG-PC	DS	Symbol
204731_at	6.43773E-05	0	0	0	X	1	MZT2B
204793_at	0.00165109	0	X	0	0	1	KDM5A
204863_s_at	6.67773E-05	0	0	0	X	1	0
204870_s_at	5.92773E-05	0	0	X	0	1	EGLN1
204957_at	4.46771E-05	0	0	X	X	2	THRAP3
204992_s_at	0.007845493	0	X	0	0	1	CTNNB1
205012_s_at	0.001394337	0	X	0	0	1	IRF2BP2
205230_at	4.47771E-05	0	0	X	X	2	LIFR
205257_s_at	4.48771E-05	0	0	X	X	2	TRAPPC5
205263_at	0.000364028	0	X	0	0	1	0
205273_s_at	3.71564E-06	0	X	0	0	1	0
205277_at	0.002817993	0	X	0	X	2	TAF1D
205279_s_at	9.46529E-05	0	X	0	0	1	0
205348_s_at	0.004175471	0	X	0	0	1	TAP1
205353_s_at	5.95128E-06	0	X	X	X	3	MTHFD1
205383_s_at	0.022308106	0	X	0	0	1	CRIM1
205434_s_at	6.44773E-05	0	0	0	X	1	C16orf62
205480_s_at	0.001823832	0	X	0	0	1	CRCP
205551_at	0.001663989	0	X	0	0	1	IL6ST
205570_at	4.49771E-05	0	0	X	X	2	ADD3
205594_at	5.82753E-05	0	X	X	X	3	ZNF264
205596_s_at	3.63711E-05	0	X	0	X	2	MACF1
205609_at	4.50771E-05	0	0	X	0	1	TSPAN5
205690_s_at	0.00151141	0	X	0	0	1	MED13L
205702_at	0.002410555	0	X	0	0	1	NACC2
205711_x_at	0.000239621	0	X	0	0	1	TMSB10
205816_at	0.071190119	0	X	0	X	2	UBE2W
205871_at	2.85538E-05	0	X	0	X	2	CFAP46
205882_x_at	6.68773E-05	0	0	0	X	1	MRPL36
205917_at	6.69773E-05	0	0	0	X	1	ZAK
205932_s_at	0.114904709	0	X	0	0	1	MRPL55
205967_at	4.51771E-05	0	0	X	X	2	0
206051_at	0.000922872	0	X	0	0	1	NFASC
206122_at	4.52771E-05	0	0	X	X	2	IL10RB-AS1
206169_x_at	0.001599795	0	X	0	0	1	TRAPPC2L
206273_at	4.53771E-05	0	0	X	X	2	TNFRSF10D
206275_s_at	0.002303848	0	X	0	0	1	PARD3B
206381_at	0.002034486	0	X	0	0	1	KDSR
206542_s_at	0.000994269	0	X	0	0	1	0
206547_s_at	4.54771E-05	0	0	X	X	2	COL22A1
206621_s_at	0.000721123	0	X	0	0	1	0
206652_at	0.001504566	0	X	X	X	3	GLTSCR1L
206710_s_at	0.000642318	0	X	0	0	1	AKAP8L
206809_s_at	0.004115506	0	X	0	0	1	UBE2QL1
206879_s_at	0.000434379	0	X	0	0	1	SRSF2
207081_s_at	4.89432E-05	0	X	0	0	1	0
207149_at	0.040169204	0	X	0	0	1	BB
207232_s_at	0.00033519	0	X	0	0	1	DD
207332_s_at	5.93773E-05	0	0	X	0	1	FAM178A
207358_x_at	0.092890848	0	0	0	X	1	M6PR
207598_x_at	4.55771E-05	0	0	X	X	2	CAPRIN2
207614_s_at	0.000145421	0	X	0	0	1	ATP5O
207730_x_at	0.00104954	0	X	X	0	2	GLTSCR1L
207922_s_at	0.021021267	0	X	0	0	1	0
207966_s_at	1.48543E-06	0	X	0	X	2	MORC2

Sensitivity analysis result. (continued)

Probe ID	BF-value	CABk	EC	HIP	MTG-PC	DS	Symbol
208002_s_at	0.008738972	0	X	0	0	1	RNMT
208066_s_at	0.011436694	0	X	0	0	1	TUBA1B
208073_x_at	9.92668E-05	0	X	0	0	1	GLRB
208517_x_at	0.000371209	0	X	X	0	2	PARK7
208610_s_at	0.000160455	0	X	0	0	1	0
208611_s_at	4.56771E-05	0	0	X	X	2	TTC3
208640_at	0.00380987	0	X	0	0	1	HSP90AB1
208641_s_at	0.033282706	0	X	0	0	1	PPP2R1A
208649_s_at	0.002369211	0	X	0	0	1	OCIAD1
208652_at	0.000714245	0	X	0	0	1	0
208678_at	0.002407466	0	X	0	0	1	PFKFB3
208680_at	0.000912629	0	X	0	0	1	CCT2
208682_s_at	0.000382275	0	X	0	0	1	0
208687_x_at	0.000628364	0	X	X	0	2	GDI2
208709_s_at	0.001757645	0	X	0	X	2	ZNF609
208710_s_at	4.57771E-05	0	0	X	X	2	0
208720_s_at	2.10068E-05	0	0	X	X	2	ANKRD12
208742_s_at	0.001841596	0	X	0	0	1	IMMT
208758_at	0.000481779	0	X	0	0	1	CLTA
208761_s_at	0.001750121	0	X	0	0	1	TERF2IP
208771_s_at	0.003083846	0	X	0	0	1	0
208781_x_at	0.008122365	0	X	0	0	1	SLC25A12
208786_s_at	0.003405053	0	X	0	0	1	ZBTB47
208799_at	0.001466286	0	X	0	0	1	DAAM1
208813_at	0.000736863	0	X	0	0	1	ZNF516
208827_at	0.000525531	0	X	X	0	2	C5orf24
208835_s_at	0.001270606	0	X	X	X	3	ZNF264
208846_s_at	0.001563533	0	X	0	0	1	CUL1
208859_s_at	3.74441E-05	0	X	X	X	3	0
208870_x_at	0.000449526	0	X	0	0	1	DNAJC7
208887_at	4.59771E-05	0	0	X	X	2	0
208898_at	0.0042551	0	X	0	0	1	RAB11FIP3
208909_at	0.004581432	0	X	0	0	1	0
208942_s_at	3.43776E-05	0	X	X	X	3	AGPAT3
208955_at	4.60771E-05	0	0	X	X	2	FAM63A
208972_s_at	4.61771E-05	0	0	X	X	2	TUBA1B
208977_x_at	0.000211311	0	X	0	0	1	LDHB
208988_at	0.001877448	0	X	0	0	1	DCTN2
208996_s_at	4.62771E-05	0	0	X	X	2	SRRM2
209014_at	0.00057841	0	X	0	0	1	COPS2
209026_x_at	0.000657444	0	X	0	X	2	BACE2
209066_x_at	4.63771E-05	0	0	X	X	2	GLIS1
209079_x_at	0.002004498	0	X	X	X	3	AMN1
209092_s_at	0.00027781	0	X	0	0	1	TAF3
209095_at	0.00109975	0	X	0	0	1	TCF25
209104_s_at	0.001745453	0	X	0	X	2	MAPK9
209119_x_at	0.010569835	0	X	0	0	1	TUBB4B
209159_s_at	1.12874E-05	0	X	0	0	1	0
209180_at	0.001020569	0	X	0	0	1	SLC25A5
209194_at	0.005470081	0	X	0	0	1	GSTA4
209200_at	0.082404147	0	X	0	0	1	0
209214_s_at	0.011035237	0	X	0	0	1	CHERP
209224_s_at	0.003638801	0	X	0	0	1	SCAF4
209225_x_at	0.000307566	0	X	X	0	2	0
209234_at	0.001705607	0	X	0	0	1	0

Sensitivity analysis result. (continued)

Probe ID	BF-value	CABk	EC	HIP	MTG-PC	DS	Symbol
209248_at	0.000922045	0	X	0	0	1	0
209249_s_at	5.94773E-05	0	0	X	0	1	KLC1
209251_x_at	4.64771E-05	0	0	X	X	2	UBE2Z
209265_s_at	0.00507009	0	X	X	0	2	FAM161B
209268_at	4.65771E-05	0	0	X	X	2	CRMP1
209276_s_at	0.002333898	0	X	0	0	1	0
209289_at	4.66771E-05	0	0	X	X	2	YPEL5
209343_at	4.67771E-05	0	0	X	X	2	TUBB4B
209372_x_at	0.00683938	0	X	0	0	1	ATP5C1
209394_at	5.95773E-05	0	0	X	0	1	0
209395_at	0.082497969	0	X	0	0	1	FAM131A
209409_at	4.68771E-05	0	0	X	X	2	0
209440_at	0.006883089	0	X	X	0	2	0
209445_x_at	0.000566899	0	X	0	0	1	0
209479_at	5.96773E-05	0	0	X	0	1	SPIRE2
209484_s_at	1.65758E-05	0	X	0	X	2	OSBPL7
209503_s_at	0.000253944	0	X	0	0	1	0
209534_x_at	0.001519661	0	X	X	X	3	TSPAN7
209537_at	0.000255808	0	X	0	0	1	ATP5C1
209553_at	0.005278964	0	X	X	X	3	WIZ
209558_s_at	4.69771E-05	0	0	X	X	2	TCF4
209569_x_at	0.006643396	0	X	0	0	1	QKI
209570_s_at	0.080009209	0	X	0	0	1	HSPA8
209586_s_at	4.70771E-05	0	0	X	X	2	PSMC5
209609_s_at	4.71771E-05	0	0	X	X	2	E
209686_at	4.72771E-05	0	0	X	X	2	C11orf73
209715_at	4.73771E-05	0	0	X	X	2	GLYR1
209751_s_at	6.45773E-05	0	0	0	X	1	0
209755_at	0.000402096	0	X	0	0	1	KATNBL1
209814_at	4.74771E-05	0	0	X	X	2	MAPT
209818_s_at	0.005434978	0	X	0	0	1	UBE2W
209839_at	0.002683556	0	X	0	0	1	CDC123
209890_at	6.71773E-05	0	0	0	X	1	CASP6
209902_at	3.58256E-05	0	X	0	0	1	SPPL3
209991_x_at	4.75771E-05	0	0	X	X	2	SECISBP2L
210094_s_at	0.152593529	0	X	0	0	1	GLOD4
210111_s_at	4.76771E-05	0	0	X	X	2	0
210156_s_at	0.032136244	0	X	0	0	1	ENC1
210252_s_at	0.001158074	0	X	0	0	1	TSPAN5
210338_s_at	0.003882499	0	X	0	0	1	TMBIM4
210501_x_at	0.001167914	0	X	0	0	1	0
210532_s_at	0.00078602	0	X	0	0	1	MKNK1
210574_s_at	0.056479929	0	X	0	X	2	ZNF785
210679_x_at	4.77771E-05	0	0	X	X	2	ZNF37A
210701_at	6.46773E-05	0	0	0	X	1	DCUN1D1
210715_s_at	0.000685307	0	X	0	0	1	TNPO1
210736_x_at	0.022046999	0	X	0	0	1	RBFO
210759_s_at	0.000805105	0	X	0	0	1	AJAP1
210766_s_at	0.000935716	0	X	0	0	1	COMMD3
210825_s_at	1.46585E-05	0	X	0	X	2	ZNF785
210840_s_at	4.78771E-05	0	0	X	X	2	DZIP3
210949_s_at	0.000305096	0	X	0	0	1	DVL1
211025_x_at	0.00079375	0	X	0	0	1	0
211270_x_at	0.057049672	0	X	0	0	1	DOCK3
211297_s_at	0.005409069	0	X	0	0	1	MARS

Sensitivity analysis result. (continued)

Probe ID	BF-value	CABk	EC	HIP	MTG-PC	DS	Symbol
211370_s_at	0.040235021	0	X	0	0	1	ABCC10
211452_x_at	0.0008002	0	X	X	0	2	CYFIP2
211464_x_at	0.000271971	0	X	X	0	2	0
211616_s_at	0.033536575	0	X	0	X	2	PRF1
211662_s_at	0.004523486	0	X	0	0	1	SLC7A14
211685_s_at	0.000659445	0	X	0	0	1	VSNL1
211779_x_at	4.79771E-05	0	0	X	X	2	0
211928_at	0.00043023	0	X	0	0	1	EIF5B
211941_s_at	0.00062192	0	X	0	0	1	0
211964_at	4.80771E-05	0	0	X	X	2	FAM162A
212017_at	0.000777055	0	X	0	X	2	BCL10
212027_at	0.004202235	0	X	X	0	2	BTF3
212037_at	4.81771E-05	0	0	X	X	2	UBN2
212062_at	0.003390467	0	X	0	0	1	ZBTB7A
212088_at	4.82771E-05	0	0	X	X	2	SCAMP2
212089_at	0.01077	0	X	X	X	3	TAP1
212095_s_at	4.83771E-05	0	0	X	X	2	HSPA12A
212104_s_at	4.84771E-05	0	0	X	X	2	MAGED2
212111_at	0.001251446	0	X	0	0	1	LYST
212155_at	0.004466281	0	X	0	0	1	0
212159_x_at	0.000489682	0	X	0	0	1	ANKRD13D
212176_at	5.97773E-05	0	0	X	X	2	NMNAT2
212177_at	9.52673E-06	0	X	0	X	2	RBFO
212208_at	6.72773E-05	0	0	0	X	1	KLF7
212209_at	4.85771E-05	0	0	X	X	2	MAPKBP1
212214_at	0.001542364	0	X	0	0	1	HABP4
212228_s_at	4.86771E-05	0	0	X	X	2	CCDC28A
212242_at	0.00189531	0	X	0	0	1	DYNC1H1
212265_at	0.00024739	0	X	0	X	2	AKAP13
212270_x_at	0.003951755	0	X	0	0	1	NRG2
212277_at	1.87529E-05	0	X	0	0	1	CPSF3
212322_at	0.007436574	0	X	0	X	2	ATP5C1
212372_at	4.87771E-05	0	0	X	X	2	ORAI2
212377_s_at	5.98773E-05	0	0	X	X	2	U2AF1
212386_at	0.003954284	0	X	0	0	1	0
212411_at	0.042096158	0	X	0	0	1	CO
212432_at	4.88772E-05	0	0	X	X	2	CAV2
212451_at	0.00027562	0	X	0	X	2	TMEM43
212468_at	6.47773E-05	0	0	0	X	1	ATIC
212492_s_at	0.000494176	0	X	0	0	1	IRF2BP2
212508_at	0.013755063	0	X	0	0	1	RECQL
212532_s_at	0.002449454	0	X	0	0	1	PIDD1
212537_x_at	4.89772E-05	0	0	X	0	1	AP2A2
212551_at	0.007356783	0	X	0	0	1	0
212614_at	5.99773E-05	0	0	X	0	1	KDM4B
212620_at	0.000119896	0	X	0	0	1	SERP2
212632_at	0.001651256	0	X	0	0	1	0
212639_x_at	9.26307E-05	0	X	X	0	2	WDR7
212652_s_at	4.90772E-05	0	0	X	X	2	PNN
212674_s_at	0.000608376	0	X	0	0	1	0
212696_s_at	0.022801945	0	X	0	0	1	PSMB6
212699_at	0.004151678	0	X	0	0	1	DNAJC1
212706_at	0.013808392	0	X	0	0	1	EIF3K
212707_s_at	0.031919953	0	X	0	0	1	SPOCK1
212716_s_at	6.00773E-05	0	0	X	0	1	GLS

Sensitivity analysis result. (continued)

Probe ID	BF-value	CABk	EC	HIP	MTG-PC	DS	Symbol
212717_at	0.001069842	0	X	X	0	2	ITFG1
212727_at	0.009920495	0	X	X	X	3	CDC16
212739_s_at	0.007927749	0	X	0	0	1	LOC100129447
212787_at	0.001530584	0	X	0	0	1	NLN
212798_s_at	0.002917192	0	X	0	0	1	COA1
212852_s_at	4.91772E-05	0	0	X	X	2	MRPL18
212877_at	0.000194392	0	X	0	0	1	POP4
212878_s_at	0.000791156	0	X	0	0	1	YTHDF2
212880_at	0.000502399	0	X	0	0	1	PHF6
212881_at	0.009164655	0	X	0	0	1	MAGED1
212896_at	0.007768489	0	X	X	X	3	KDM6B
212904_at	0.002174939	0	X	0	0	1	CRBN
212917_x_at	0.000484422	0	X	0	0	1	PHB
212922_s_at	0.035986805	0	X	0	0	1	0
212923_s_at	6.01773E-05	0	0	X	0	1	RAN
212959_s_at	0.019406584	0	X	0	X	2	CNOT10
212964_at	0.001671251	0	X	0	0	1	NMT2
212993_at	6.73773E-05	0	0	0	X	1	TLE4
212995_x_at	6.48773E-05	0	0	0	X	1	CAPRIN1
213009_s_at	0.005298258	0	X	0	0	1	GOLGA7
213032_at	4.92772E-05	0	0	X	X	2	FASTK
213079_at	0.006370068	0	X	0	X	2	TMEM106A
213089_at	4.93772E-05	0	0	X	X	2	HSPA8
213137_s_at	0.001643449	0	X	X	0	2	DNAJA4
213195_at	4.94772E-05	0	0	X	X	2	EPB41L3
213203_at	0.008092317	0	X	0	0	1	SAV1
213268_at	0.002331459	0	X	0	0	1	TUBB
213275_x_at	6.02773E-05	0	0	X	0	1	0
213328_at	0.001828157	0	X	0	0	1	NCALD
213333_at	4.95316E-05	0	X	0	0	1	TFRC
213366_x_at	0.000213121	0	X	0	0	1	VPS33B
213386_at	0.001167271	0	X	0	0	1	HMO
213394_at	0.000415899	0	X	0	0	1	GHITM
213400_s_at	0.04226455	0	X	0	0	1	0
213411_at	0.001050399	0	X	0	0	1	WARS
213421_x_at	0.022845314	0	X	X	X	3	COA3
213476_x_at	0.003236457	0	X	X	0	2	0
213482_at	0.000337569	0	X	0	0	1	SPINT2
213484_at	0.005000029	0	X	0	0	1	ORMDL1
213485_s_at	0.000346675	0	X	0	0	1	LINC00662
213530_at	0.005181297	0	X	0	X	2	MRPS21
213535_s_at	4.95772E-05	0	0	X	0	1	0
213545_x_at	4.96772E-05	0	0	X	X	2	FAHD2A
213636_at	0.000870665	0	X	0	0	1	KIAA0513
213668_s_at	4.97772E-05	0	0	X	0	1	C5orf22
213682_at	0.013760878	0	X	0	0	1	PPP2CA
213693_s_at	0.008777255	0	X	X	X	3	RNF8
213726_x_at	0.000179258	0	X	X	0	2	0
213735_s_at	0.003677657	0	X	0	X	2	DHCR24
213744_at	4.98772E-05	0	0	X	X	2	EIF4H
213808_at	0.019283615	0	X	0	0	1	PRD
213938_at	0.00344531	0	X	0	0	1	0
214043_at	4.99772E-05	0	0	X	X	2	GOT1
214075_at	0.003278754	0	X	0	0	1	LOC144438
214086_s_at	0.006852063	0	X	0	0	1	USP42

Sensitivity analysis result. (continued)

Probe ID	BF-value	CABk	EC	HIP	MTG-PC	DS	Symbol
214114_x_at	0.000621391	0	X	0	0	1	CNRIP1
214170_x_at	0.001769456	0	X	0	0	1	FAM168B
214250_at	0.019934469	0	X	0	0	1	CLIP1
214353_at	0.192078149	0	X	0	0	1	ENO2
214359_s_at	0.00093349	0	X	0	0	1	C14orf2
214394_x_at	5.00772E-05	0	0	X	X	2	MZT2B
214402_s_at	0.065707452	0	X	0	0	1	KLC1
214415_at	0.000582216	0	X	0	0	1	CO
214432_at	5.01772E-05	0	0	X	0	1	ZNF148
214434_at	0.000379654	0	X	0	0	1	LRRFIP1
214439_x_at	0.013245406	0	X	X	X	3	PSMA1
214501_s_at	6.74773E-05	0	0	0	X	1	BTN2A1
214623_at	5.02772E-05	0	0	X	0	1	DCTN6
214659_x_at	4.98402E-06	0	X	0	0	1	UB
214743_at	5.03772E-05	0	0	X	X	2	ZFR
214761_at	6.03773E-05	0	0	X	0	1	MAZ
214782_at	6.1898E-05	0	X	0	0	1	TNPO1
214799_at	6.83506E-05	0	X	X	X	3	PSMD1
214815_at	0.012580962	0	X	0	0	1	KIAA1045
214821_at	0.002763253	0	X	0	0	1	CLTA
214823_at	0.003828683	0	X	X	0	2	0
214850_at	5.04772E-05	0	0	X	X	2	0
214864_s_at	6.4869E-05	0	X	0	0	1	TPRG1L
214924_s_at	0.00473153	0	X	0	0	1	NDUFS8
214925_s_at	0.012452219	0	X	X	0	2	PRD
214933_at	0.004442789	0	X	0	0	1	0
214934_at	5.05772E-05	0	0	X	X	2	GHITM
215019_x_at	0.005957041	0	X	0	0	1	ELAVL4
215020_at	0.005079055	0	X	0	0	1	NIPSNAP1
215069_at	0.000599286	0	X	0	0	1	HSP90AB1
215169_at	0.001508464	0	X	0	0	1	SF3B5
215191_at	0.007882723	0	X	0	0	1	0
215230_x_at	4.79339E-05	0	X	0	0	1	CSE1L
215268_at	0.017358279	0	X	0	0	1	C14orf166
215269_at	6.04773E-05	0	0	X	0	1	PARP11
215306_at	5.06772E-05	0	0	X	X	2	SN
215344_at	0.004780847	0	X	X	0	2	FB
215373_x_at	0.003343507	0	X	0	0	1	PITPNA
215499_at	6.05773E-05	0	0	X	0	1	AMBRA1
215504_x_at	5.07772E-05	0	0	X	X	2	ASH1L
215514_at	5.08772E-05	0	0	X	X	2	PSMC3
215553_x_at	5.09772E-05	0	0	X	X	2	MRPS23
215587_x_at	0.011136704	0	X	0	0	1	SMARCA2
215600_x_at	5.10772E-05	0	0	X	X	2	USP53
215693_x_at	6.06773E-05	0	0	X	0	1	RABGGTB
215698_at	6.49773E-05	0	0	0	X	1	SAFB2
215764_x_at	0.023467832	0	X	X	X	3	UCHL1
215789_s_at	0.000326654	0	X	0	0	1	NELL2
215889_at	0.050670244	0	X	X	X	3	0
215908_at	0.027970611	0	X	0	0	1	C11orf96
215963_x_at	0.025971587	0	X	X	X	3	0
215978_x_at	0.001439506	0	X	0	0	1	0
216187_x_at	0.001449724	0	X	0	0	1	0
216210_x_at	0.022812275	0	X	0	0	1	PLEKHM1
216295_s_at	0.000878018	0	X	0	0	1	AKR1B1

Sensitivity analysis result. (continued)

Probe ID	BF-value	CABk	EC	HIP	MTG-PC	DS	Symbol
216308_x_at	0.010140246	0	X	X	X	3	CHD9
216384_x_at	5.11772E-05	0	0	X	0	1	DLD
216449_x_at	0.009267563	0	X	0	0	1	SND1
216508_x_at	6.75773E-05	0	0	0	X	1	ZNF562
216524_x_at	5.12772E-05	0	0	X	X	2	CSNK2B
216550_x_at	0.000120274	0	X	X	X	3	IQCA1
216903_s_at	0.001393051	0	X	0	0	1	LOC100289333
216977_x_at	0.004125457	0	X	0	0	1	KIZ
217152_at	0.007025512	0	X	0	0	1	INA
217164_at	0.004025891	0	X	X	0	2	APOPT1
217446_x_at	0.00014893	0	X	0	0	1	FH
217457_s_at	0.001458213	0	X	0	0	1	MADD
217482_at	6.12112E-06	0	X	0	X	2	BAALC
217554_at	0.000500399	0	X	0	X	2	UQCRC2
217579_x_at	0.002898486	0	X	0	0	1	0
217679_x_at	0.001622954	0	X	0	0	1	TMEM246
217703_x_at	4.9451E-05	0	X	0	0	1	EIF3K
217721_at	0.012992586	0	X	X	X	3	CNTRL
217727_x_at	0.041917581	0	X	0	0	1	0
217731_s_at	0.002376542	0	X	0	0	1	NPAS3
217733_s_at	6.76773E-05	0	0	0	X	1	PPARD
217759_at	5.13772E-05	0	0	X	0	1	0
217768_at	0.000945808	0	X	0	0	1	CSNK1G2
217773_s_at	0.002120355	0	X	0	0	1	NBAS
217802_s_at	0.021538432	0	X	0	0	1	ST
217819_at	0.000616315	0	X	X	0	2	0
217841_s_at	0.017170098	0	X	0	0	1	CTSB
217860_at	4.57924E-07	0	X	X	X	3	LUC7L3
217867_x_at	0.000161568	0	X	0	0	1	SN
217883_at	5.26796E-05	0	X	0	0	1	CO
217897_at	6.07773E-05	0	0	X	0	1	NAE1
217900_at	0.002525111	0	X	0	0	1	FAM162A
217904_s_at	5.14772E-05	0	0	X	X	2	LOC728153
217907_at	0.000570688	0	X	0	0	1	LOC222070
217927_at	0.001379676	0	X	X	0	2	SPCS1
217937_s_at	0.03638152	0	X	0	X	2	MICU1
217939_s_at	5.15772E-05	0	0	X	0	1	YY1AP1
217946_s_at	5.16772E-05	0	0	X	X	2	CHMP2A
217969_at	5.17772E-05	0	0	X	X	2	0
217985_s_at	0.005939217	0	X	0	0	1	TBK1
218009_s_at	6.08773E-05	0	0	X	0	1	DYNLL1
218026_at	0.000681126	0	X	X	X	3	MCTS1
218027_at	0.014165434	0	X	0	0	1	PAFAH1B1
218048_at	0.000331747	0	X	X	X	3	ZNF721
218097_s_at	0.012682091	0	X	0	0	1	0
218101_s_at	0.00889948	0	X	0	0	1	FAM162A
218120_s_at	0.000667517	0	X	0	0	1	0
218143_s_at	0.000376515	0	X	0	0	1	0
218163_at	0.00142653	0	X	0	0	1	RAP1GDS1
218190_s_at	0.000199834	0	X	0	0	1	TNPO2
218220_at	6.09773E-05	0	0	X	0	1	PSMB5
218241_at	0.01051735	0	X	0	0	1	GABRD
218247_s_at	5.18772E-05	0	0	X	X	2	SLC35E2
218271_s_at	0.042628432	0	X	X	X	3	BUD31
218298_s_at	5.19772E-05	0	0	X	X	2	0

Sensitivity analysis result. (continued)

Probe ID	BF-value	CABk	EC	HIP	MTG-PC	DS	Symbol
218302_at	5.20772E-05	0	0	X	X	2	YLPM1
218310_at	0.002747315	0	X	0	0	1	TRAPPC10
218330_s_at	5.66824E-06	0	X	X	X	3	PRKCZ
218354_at	6.84596E-05	0	X	0	0	1	0
218415_at	0.000661328	0	X	0	0	1	OPA1
218418_s_at	6.10773E-05	0	0	X	0	1	VDAC3
218455_at	0.006185296	0	X	0	0	1	LDHA
218456_at	8.27892E-05	0	X	0	0	1	ZC3H7B
218491_s_at	0.003605315	0	X	0	0	1	0
218504_at	0.000705895	0	X	0	0	1	COPS5
218520_at	0.001417213	0	X	X	X	3	TSPAN31
218521_s_at	8.69569E-06	0	0	0	X	1	0
218623_at	0.028083369	0	X	0	0	1	PTPN2
218625_at	0.004727	0	X	X	0	2	GPRASP1
218628_at	0.004612144	0	X	0	0	1	ST
218642_s_at	0.04224308	0	X	0	0	1	SRRT
218843_at	5.21772E-05	0	0	X	X	2	SV2B
218865_at	5.22772E-05	0	0	X	X	2	RPS6KC1
218909_at	0.001664642	0	X	0	0	1	HIC2
218980_at	0.002188538	0	X	0	0	1	KIF1B
219007_at	6.11773E-05	0	0	X	0	1	ZNF777
219019_at	0.000486338	0	X	X	0	2	NHP2
219032_x_at	5.23772E-05	0	0	X	X	2	SUMO1
219108_x_at	6.50773E-05	0	0	0	X	1	0
219163_at	0.001102545	0	X	0	0	1	NRD1
219203_at	5.24772E-05	0	0	X	X	2	FH
219206_x_at	0.000291524	0	X	0	0	1	GFM2
219342_at	6.12566E-05	0	X	0	0	1	CCAR1
219356_s_at	0.039821049	0	X	0	0	1	PSMB7
219425_at	0.002080659	0	X	0	0	1	ANKRD11
219549_s_at	5.68054E-06	0	X	0	0	1	0
219553_at	0.132164804	0	X	0	0	1	UGP2
219569_s_at	0.003630174	0	X	0	0	1	NEK1
219671_at	5.25772E-05	0	0	X	X	2	CFAP46
219709_x_at	5.26772E-05	0	0	X	X	2	GALNT15
219896_at	0.001437818	0	X	0	0	1	ARHGEF1
219945_at	7.86464E-05	0	X	0	0	1	SAP18
219961_s_at	0.001129177	0	X	0	X	2	CCNDBP1
220071_x_at	5.27772E-05	0	0	X	X	2	0
220136_s_at	5.28772E-05	0	0	X	X	2	KDM2A
220269_at	5.29772E-05	0	0	X	X	2	SET
220295_x_at	5.30772E-05	0	0	X	X	2	0
220316_at	5.31772E-05	0	0	X	0	1	TUBA4A
220411_x_at	0.004195306	0	X	0	0	1	0
220539_at	6.78773E-05	0	0	0	X	1	OCIAD1
220609_at	0.003174766	0	X	X	X	3	FKSG49
220615_s_at	5.32772E-05	0	0	X	X	2	STAU2
220720_x_at	0.000788859	0	X	0	0	1	ATP5B
220864_s_at	0.002606537	0	X	0	0	1	0
220942_x_at	0.001439766	0	X	X	0	2	PSMB1
220948_s_at	4.76098E-05	0	X	0	0	1	0
220966_x_at	0.000264668	0	X	X	X	3	DCHS1
220999_s_at	0.00035083	0	X	0	0	1	SCN2A
221012_s_at	0.003814908	0	X	0	0	1	0
221155_x_at	0.000336517	0	X	0	0	1	SO

Sensitivity analysis result. (continued)

Probe ID	BF-value	CABk	EC	HIP	MTG-PC	DS	Symbol
221263_s_at	0.000934349	0	X	0	0	1	RPS16P5
221449_s_at	0.00053794	0	X	0	0	1	FBRSL1
221486_at	5.33772E-05	0	0	X	X	2	ACTN2
221488_s_at	5.34772E-05	0	0	X	X	2	PSMG3
221495_s_at	0.000174412	0	X	0	0	1	PITHD1
221497_x_at	6.51773E-05	0	0	0	X	1	NDUFA4
221506_s_at	0.001463852	0	X	X	X	3	ADAM33
221510_s_at	0.014318017	0	X	0	0	1	PSMA5
221526_x_at	5.35772E-05	0	0	X	0	1	SLC25A29
221540_x_at	0.000465046	0	X	0	0	1	0
221573_at	0.013835146	0	X	0	0	1	LRRC47
221688_s_at	0.044431583	0	X	0	0	1	FHOD3
221702_s_at	0.000937614	0	X	X	0	2	KLF15
221711_s_at	0.007772311	0	X	0	0	1	0
221772_s_at	5.36772E-05	0	0	X	X	2	0
221829_s_at	0.000848475	0	X	0	0	1	0
221864_at	5.37772E-05	0	0	X	X	2	DGUOK
221877_at	0.011335524	0	X	X	X	3	NDUFA1
221942_s_at	5.38772E-05	0	0	X	X	2	ME3
222024_s_at	0.000433703	0	X	X	X	3	MALAT1
222043_at	0.030270173	0	X	X	X	3	EHMT1
222047_s_at	0.001651837	0	X	X	X	3	CHGB
222052_at	0.00273274	0	X	X	0	2	SIK3
222101_s_at	0.002033271	0	X	0	0	1	0
222113_s_at	0.003609316	0	X	0	0	1	MNT
222126_at	6.12773E-05	0	0	X	0	1	NHP2L1
222160_at	7.00148E-05	0	X	X	X	3	CAMTA1
222282_at	0.001808889	0	X	0	0	1	0
222284_at	5.39772E-05	0	0	X	X	2	VCP
222294_s_at	5.40772E-05	0	0	X	X	2	ITM2B
222310_at	0.000192199	0	X	0	0	1	ATP6V1E1
222313_at	2.01345E-05	0	X	0	X	2	PHTF1
222339_x_at	0.002559879	0	X	0	0	1	0
222368_at	0.047980638	0	X	0	0	1	RPL3
222395_s_at	0.000194408	0	X	0	0	1	RPL17
222408_s_at	0.000207744	0	X	0	0	1	AN
222418_s_at	0.000472067	0	X	0	0	1	LSM12
222430_s_at	0.000576248	0	X	0	0	1	DNASE1
222439_s_at	6.52773E-05	0	0	0	X	1	0
222446_s_at	6.13773E-05	0	0	X	X	2	IARS2
222457_s_at	5.41772E-05	0	0	X	X	2	UBE2I
222533_at	0.000579844	0	X	0	0	1	0
222605_at	0.003270421	0	X	0	0	1	TET2
222610_s_at	6.53773E-05	0	0	0	X	1	TUBB3
222620_s_at	0.000528739	0	X	0	0	1	0
222621_at	0.014396476	0	X	0	0	1	0
222652_s_at	0.000259679	0	X	0	0	1	WDR81
222657_s_at	0.000270044	0	X	0	0	1	IMPDH2
222669_s_at	0.036134773	0	X	0	0	1	DNM3
222728_s_at	6.61738E-05	0	X	X	0	2	FNIP1
222745_s_at	5.42772E-05	0	0	X	X	2	CASC10
222780_s_at	0.001159037	0	X	0	0	1	C19orf54
222786_at	0.012703689	0	X	0	0	1	RABGEF1
222976_s_at	0.003844551	0	X	0	0	1	PTPRM
222984_at	0.002859966	0	X	0	0	1	SLC25A4

Sensitivity analysis result. (continued)

Probe ID	BF-value	CABk	EC	HIP	MTG-PC	DS	Symbol
222988_s_at	0.006646583	0	X	0	0	1	RRAGA
222993_at	6.14773E-05	0	0	X	0	1	0
222994_at	0.000733116	0	X	0	0	1	CEP295
222997_s_at	0.000696385	0	X	0	0	1	C1orf61
223004_s_at	0.004247638	0	X	0	0	1	PRDM2
223011_s_at	0.000107279	0	X	0	0	1	FAM20B
223026_s_at	0.010608744	0	X	0	0	1	ZNF618
223043_at	0.018488943	0	X	0	0	1	NDUFS3
223053_x_at	0.027834185	0	X	0	0	1	PAIP2
223084_s_at	0.001876832	0	X	X	0	2	STMN2
223091_x_at	0.00954412	0	X	0	0	1	0
223124_s_at	0.002108533	0	X	0	0	1	SDHB
223134_at	7.83555E-05	0	X	X	X	3	PSMB4
223156_at	0.000993234	0	X	0	0	1	ANKMY2
223164_at	0.034722451	0	X	0	0	1	ARHGAP21
223174_at	0.006492216	0	X	0	0	1	LOC728730
223183_at	0.0001502	0	X	0	0	1	EMC3
223185_s_at	5.43772E-05	0	0	X	X	2	0
223187_s_at	0.000688797	0	X	0	0	1	SCD
223193_x_at	0.000363106	0	X	0	0	1	LATS2
223215_s_at	0.001332494	0	X	0	0	1	ZNRF1
223219_s_at	0.000591808	0	X	0	0	1	LTA4H
223363_at	0.002106402	0	X	0	0	1	AARS
223380_s_at	0.003044464	0	X	X	X	3	BMP2K
223460_at	0.000334245	0	X	0	0	1	F
223672_at	0.006666156	0	X	0	0	1	ENGASE
223673_at	5.44772E-05	0	0	X	X	2	LOC202181
223679_at	6.54773E-05	0	0	0	X	1	0
223748_at	0.061358029	0	X	0	0	1	0
223857_x_at	0.001119565	0	X	0	0	1	RGS4
223940_x_at	6.55773E-05	0	0	0	X	1	TUBB3
224076_s_at	0.006891993	0	X	0	0	1	RCOR3
224151_s_at	5.45772E-05	0	0	X	X	2	SRP54
224163_s_at	0.000530007	0	X	0	0	1	NENF
224187_x_at	0.000252194	0	X	0	0	1	NAV1
224288_x_at	0.001939099	0	X	0	0	1	NFIC
224331_s_at	6.79773E-05	0	0	0	X	1	0
224345_x_at	0.001353117	0	X	0	0	1	RBFA
224360_s_at	6.15773E-05	0	0	X	0	1	0
224452_s_at	4.74344E-05	0	X	0	0	1	EIF2B1
224514_x_at	0.279121938	0	X	0	0	1	ATP9A
224517_at	0.009052859	0	X	0	0	1	MAP1LC3B
224568_x_at	5.46772E-05	0	0	X	X	2	S100PBP
224569_s_at	6.56773E-05	0	0	0	X	1	ERC2
224597_at	2.18603E-06	0	X	0	0	1	RLIM
224613_s_at	5.47772E-05	0	0	X	X	2	0
224628_at	5.48772E-05	0	0	X	X	2	ERCC1
224636_at	0.003674358	0	X	0	0	1	JUND
224736_at	0.001796791	0	X	X	X	3	0
224737_x_at	5.49772E-05	0	0	X	X	2	0
224771_at	8.2845E-06	0	X	0	0	1	THYN1
224774_s_at	0.003291437	0	X	0	X	2	ME
224813_at	0.029710692	0	X	0	0	1	EPS15L1
224862_at	5.50772E-05	0	0	X	X	2	0
224871_at	0.000889522	0	X	0	0	1	0

Sensitivity analysis result. (continued)

Probe ID	BF-value	CABk	EC	HIP	MTG-PC	DS	Symbol
224888_at	0.004340238	0	X	0	0	1	NDUFA2
224964_s_at	0.03899698	0	X	0	0	1	WAS
225003_at	5.51772E-05	0	0	X	X	2	ZFP91
225030_at	0.006452615	0	X	0	0	1	CO
225055_at	5.52772E-05	0	0	X	X	2	RAC1
225056_at	0.016302896	0	X	0	0	1	TRIM8
225061_at	0.000632557	0	X	0	0	1	0
225082_at	0.00043756	0	X	X	0	2	ZNF204P
225092_at	5.53772E-05	0	0	X	X	2	NR
225101_s_at	0.001283452	0	X	0	0	1	TPM3
225223_at	5.54772E-05	0	0	X	X	2	HSPA8
225230_at	6.16773E-05	0	0	X	0	1	RPL17
225260_s_at	0.004343617	0	X	0	0	1	TCF4
225293_at	0.003979559	0	X	0	0	1	COL27A1
225298_at	5.55772E-05	0	0	X	X	2	0
225330_at	0.000948413	0	X	0	X	2	SNRPA1
225358_at	0.039748658	0	X	0	0	1	IQSEC1
225379_at	0.000268546	0	X	0	0	1	CCT7
225392_at	0.001779271	0	X	0	0	1	DYNC1I1
225416_at	0.003445466	0	X	0	0	1	PODNL1
225461_at	0.002312802	0	X	0	X	2	RAB18
225463_x_at	5.56772E-05	0	0	X	X	2	TIMMDC1
225484_at	5.57772E-05	0	0	X	X	2	ATP6V1D
225491_at	0.123989172	0	X	0	0	1	EPT1
225493_at	0.000759909	0	X	X	X	3	MRPL32
225501_at	0.000576414	0	X	X	X	3	PDCL
225516_at	5.58772E-05	0	0	X	X	2	0
225557_at	2.53375E-05	0	X	0	0	1	HEATR5B
225571_at	6.57773E-05	0	0	0	X	1	SH3BP5
225592_at	0.017847252	0	X	0	0	1	CACNA1A
225636_at	5.59772E-05	0	0	X	X	2	RNF187
225655_at	5.60772E-05	0	0	X	X	2	VDAC2
225657_at	0.025639488	0	X	0	0	1	0
225662_at	6.80773E-05	0	0	0	X	1	UQCRFS1
225678_at	0.022860129	0	X	0	0	1	CCDC53
225703_at	3.18144E-05	0	X	0	0	1	PARP10
225704_at	0.002048904	0	X	X	0	2	NRN1
225772_s_at	0.020908265	0	X	0	0	1	TRAK1
225776_at	6.17773E-05	0	0	X	0	1	PDK3
225781_at	0.000176297	0	X	0	0	1	0
225898_at	0.007987351	0	X	0	0	1	FB
225908_at	0.025661267	0	X	0	0	1	SET
225934_at	0.085150956	0	X	X	0	2	IQGAP1
225941_at	0.033839706	0	X	0	0	1	ARHGEF10L
225960_at	0.003067817	0	X	0	0	1	0
226062_x_at	0.000151446	0	X	0	0	1	SHROOM1
226132_s_at	0.006889325	0	0	X	0	1	0
226176_s_at	0.000762301	0	0	X	0	1	0
226199_at	0.016142525	0	X	0	0	1	CDAN1
226201_at	7.75865E-06	0	X	0	0	1	NR
226258_at	0.000167343	0	X	0	0	1	RBBP4
226285_at	0.000611109	0	X	0	0	1	NFIA
226336_at	0.094631107	0	X	0	0	1	0
226339_at	0.009239414	0	X	0	0	1	RBMS3
226397_s_at	0.053052752	0	X	0	0	1	RAB3GAP1

Sensitivity analysis result. (continued)

Probe ID	BF-value	CABk	EC	HIP	MTG-PC	DS	Symbol
226447_at	5.66772E-05	0	0	X	0	1	VPS8
226463_at	0.047002179	0	X	0	0	1	TRIM37
226520_at	0.015183686	0	0	X	0	1	0
226528_at	0.013121094	0	X	0	0	1	NDUFAB1
226554_at	0.000373833	0	X	0	0	1	CDK7
226592_at	0.002850986	0	X	0	0	1	TNK2
226618_at	7.06429E-05	0	X	0	0	1	CETN2
226675_s_at	1.18919E-06	0	X	0	0	1	LINC00663
226738_at	0.002643656	0	X	0	0	1	REEP1
226749_at	0.005627425	0	X	0	0	1	0
226751_at	0.000766182	0	X	0	0	1	MRPS9
226800_at	0.007205795	0	X	0	0	1	PDZD2
226806_s_at	0.005163925	0	X	0	0	1	UBE2N
226853_at	0.003107933	0	X	0	0	1	SCARNA17
226895_at	0.003296454	0	X	X	X	3	BLVRA
226896_at	0.005993312	0	X	0	0	1	BAZ1A
226899_at	0.000693162	0	X	0	0	1	ZNF528
226929_at	0.000814091	0	X	0	0	1	AMIGO1
227041_at	0.019718292	0	X	X	0	2	CHCHD1
227066_at	5.74772E-05	0	0	X	X	2	0
227099_s_at	0.001037403	0	X	0	0	1	NFS1
227125_at	6.83773E-05	0	0	0	X	1	NUP133
227160_s_at	0.000149491	0	X	0	0	1	EEF1B2
227179_at	0.001941696	0	X	0	0	1	ERP29
227229_at	0.068402569	0	X	0	0	1	0
227318_at	0.000229581	0	X	0	0	1	0
227334_at	0.035289464	0	X	0	0	1	TSR2
227345_at	0.01198173	0	0	0	X	1	BOD1
227441_s_at	6.23773E-05	0	0	X	0	1	BTBD10
227512_at	0.009888553	0	X	0	0	1	0
227612_at	1.3773E-05	0	X	0	0	1	MFSD4
227946_at	0.000225713	0	X	0	0	1	IFT43
227954_at	0.041199142	0	X	0	0	1	0
227964_at	0.021660635	0	X	0	0	1	CFDP1
228038_at	6.24773E-05	0	0	X	X	2	0
228131_at	0.003479756	0	X	0	0	1	TMEM9
228262_at	0.010712906	0	X	0	0	1	SGIP1
228396_at	6.25773E-05	0	0	X	0	1	KIAA0195
228411_at	0.575714505	0	0	0	X	1	0
228516_at	0.005031402	0	X	X	0	2	MANEAL
228541_x_at	0.000117	0	X	0	0	1	WHSC1L1
228613_at	0.000149283	0	X	0	0	1	KARS
228617_at	0.09182684	0	X	0	0	1	LOC100287331
228711_at	0.000302962	0	X	0	0	1	0
228768_at	4.0306E-07	0	X	0	X	2	0
228775_at	0.002954807	0	X	0	0	1	SERPINB6
228798_x_at	0.00083638	0	X	0	0	1	0
228854_at	0.000359281	0	X	0	0	1	EFCAB7
228959_at	0.004743496	0	X	0	0	1	0
229066_at	0.060390103	0	X	X	0	2	CAP2
229312_s_at	0.009295591	0	X	0	0	1	SGPL1
229319_at	0.002334432	0	X	X	0	2	SCAF11
229350_x_at	0.004662451	0	X	0	0	1	0
229483_at	0.000188165	0	X	0	0	1	0
229586_at	0.001070994	0	X	0	0	1	LOC100130429

Sensitivity analysis result. (continued)

Probe ID	BF-value	CABk	EC	HIP	MTG-PC	DS	Symbol
229603_at	0.052818932	0	X	X	0	2	SCCPDH
229649_at	0.003844469	0	X	0	0	1	SKIV2L2
229776_at	2.13807E-05	0	X	0	0	1	BABAM1
229811_at	0.000111599	0	X	0	0	1	ZAK
229850_at	6.86773E-05	0	0	0	X	1	PFN2
229878_at	0.002777465	0	X	0	0	1	0
229949_at	0.000393374	0	X	0	0	1	NME4
230027_s_at	0.009052783	0	X	0	0	1	WDR54
230058_at	3.73395E-05	0	X	0	0	1	0
230064_at	0.000222412	0	X	0	0	1	SNAPC5
230083_at	0.000996069	0	X	0	0	1	GRSF1
230245_s_at	0.015648684	0	X	0	0	1	SN
230255_at	0.001497673	0	0	X	0	1	0
230326_s_at	0.000258726	0	X	0	X	2	G3BP1
230411_at	2.23388E-05	0	X	0	0	1	0
230528_s_at	0.000315038	0	X	0	0	1	KLHL42
230590_at	0.006542697	0	X	X	0	2	KDM5B
230651_at	3.85526E-05	0	X	0	0	1	0
230663_at	0.007557297	0	X	0	0	1	0
230820_at	0.015624671	0	X	0	0	1	ACOT7
230970_at	6.87773E-05	0	0	0	X	1	MUC1
231015_at	0.002220688	0	X	0	X	2	0
231019_x_at	0.006976918	0	X	0	0	1	NDUFC2
231212_x_at	0.017082269	0	X	0	0	1	0
231220_at	0.013764841	0	X	0	0	1	MRPL43
231577_s_at	0.024802287	0	X	0	0	1	POLR2J4
231735_s_at	0.002303558	0	X	0	0	1	PIAS4
231887_s_at	0.116993215	0	X	0	0	1	SFPQ
231999_at	0.001800693	0	X	0	0	1	TRUB1
232002_at	0.002279999	0	X	0	0	1	HSP90B1
232169_x_at	0.000904728	0	X	0	0	1	GKAP1
232174_at	0.000718354	0	X	0	0	1	HA
232207_at	0.102256052	0	X	0	0	1	0
232264_at	2.0764E-07	0	X	0	0	1	TGFBR3
232280_at	0.00216634	0	0	X	X	2	B3GAT2
232333_at	0.650817849	0	X	0	0	1	ATP2B2
232341_x_at	0.000423249	0	X	0	0	1	MFF
232347_x_at	0.000361415	0	X	0	0	1	0
232511_at	5.65431E-05	0	X	0	0	1	UTP18
232516_x_at	0.001393371	0	X	0	0	1	ME
232547_at	0.040084001	0	X	0	0	1	DLG3
232597_x_at	0.007447887	0	X	0	0	1	SN
232626_at	0.00332473	0	X	0	0	1	SRSF11
232795_at	0.040770556	0	X	0	0	1	RPS7
232807_at	0.000220353	0	X	0	0	1	DYNC2LI1
232814_x_at	0.001140667	0	X	0	0	1	TRIM44
232978_at	0.008791891	0	X	0	0	1	GRHPR
233025_at	0.005699862	0	X	0	0	1	PSMB4
233273_at	0.024449528	0	X	0	0	1	GOLGA5
233359_at	0.000736486	0	X	0	0	1	NR2F2
233393_at	0.008340359	0	X	0	0	1	CAPNS1
233449_at	0.012609834	0	X	0	0	1	VPS29
233506_at	0.081693166	0	X	0	0	1	RPL4
233611_at	0.010982546	0	X	0	0	1	MAP7D2
233642_s_at	0.004425316	0	X	0	0	1	LMNA

Sensitivity analysis result. (continued)

Probe ID	BF-value	CABk	EC	HIP	MTG-PC	DS	Symbol
233775_x_at	0.001118146	0	X	0	0	1	HBP1
233868_x_at	0.002148528	0	X	0	0	1	KCNE4
234047_at	0.009312395	0	X	0	0	1	RBM
234107_s_at	0.012397833	0	X	0	0	1	0
234491_s_at	0.000653057	0	X	0	0	1	EWSR1
234505_at	1.82692E-06	0	X	0	0	1	0
234562_x_at	0.007263342	0	X	X	0	2	EGLN1
234563_at	6.27773E-05	0	0	X	0	1	0
234762_x_at	0.000544051	0	X	0	0	1	RARS
234788_x_at	8.92113E-05	0	X	0	0	1	IRGQ
235031_at	6.28773E-05	0	0	X	0	1	USP14
235070_at	0.000321731	0	X	0	0	1	GTF2B
235079_at	0.018968677	0	X	0	0	1	NRGN
235112_at	0.032910049	0	X	0	0	1	TBCB
235154_at	0.000173212	0	X	0	0	1	RPS3A
235327_x_at	0.00083268	0	X	0	0	1	0
235461_at	5.76772E-05	0	0	X	X	2	EML3
235484_at	3.06889E-05	0	X	0	0	1	0
235601_at	0.027831135	0	X	0	0	1	PMPCA
235765_at	0.000599663	0	X	0	0	1	DTD1
235850_at	0.002260586	0	X	0	0	1	SPTAN1
235954_at	7.74554E-05	0	X	0	0	1	TRIM33
235990_at	0.022369604	0	X	0	0	1	0
236007_at	0.023903995	0	X	X	0	2	CUEDC2
236240_at	0.017088719	0	X	0	0	1	CHST12
236439_at	0.006063144	0	X	0	0	1	0
236462_at	0.007201953	0	X	0	0	1	0
236524_at	0.017468647	0	X	0	0	1	0
236571_at	0.003815687	0	X	0	0	1	MT
236610_at	0.006567689	0	X	0	0	1	0
236645_at	0.010874684	0	X	0	0	1	CLCN7
236653_at	0.000690824	0	X	0	0	1	BIN1
236869_at	0.000523243	0	X	0	0	1	0
237100_at	0.00241317	0	X	0	0	1	THRAP3
237110_at	0.000207016	0	X	0	0	1	MPP1
237483_at	0.012752604	0	X	0	0	1	MOAP1
237747_at	0.011678634	0	X	0	0	1	NUP50
237868_x_at	0.000222964	0	X	0	0	1	CELF4
238134_at	0.004944178	0	X	0	0	1	0
238299_at	0.002166449	0	X	0	0	1	0
238350_at	0.000373791	0	X	0	0	1	MRPL15
238447_at	0.005178654	0	X	0	0	1	GLS
238466_at	0.0253296	0	X	0	0	1	DNAJC1
238549_at	0.024741988	0	X	0	0	1	LOC439911
238584_at	0.001114656	0	X	0	0	1	C10orf12
238589_s_at	0.001439529	0	X	0	0	1	DTYMK
238595_at	0.000703448	0	X	0	0	1	LCOR
238642_at	0.000395922	0	X	0	0	1	C14orf159
238651_at	1.10089E-05	0	X	0	0	1	0
238711_s_at	0.000794443	0	X	0	X	2	0
238718_at	0.003175659	0	X	0	0	1	LINC00926
238902_at	0.001611512	0	X	0	0	1	LOC100288418
239071_at	0.005113398	0	X	0	0	1	USP7
239144_at	0.009382296	0	0	X	0	1	UPRT
239311_at	0.001232607	0	X	0	0	1	0

Sensitivity analysis result. (continued)

Probe ID	BF-value	CABk	EC	HIP	MTG-PC	DS	Symbol
239435_x_at	0.004977249	0	X	0	0	1	0
239449_at	0.091603141	0	X	0	0	1	MRPL55
239461_at	6.30773E-05	0	0	X	0	1	0
239597_at	0.001538673	0	X	0	0	1	C4orf29
239661_at	0.004550879	0	X	0	X	2	PPME1
239748_x_at	0.001938848	0	X	0	0	1	KIAA0754
239811_at	0.00757115	0	X	0	0	1	0
239826_at	0.001876992	0	X	0	0	1	NRM
239889_at	0.000494989	0	X	0	0	1	0
240121_x_at	0.000194382	0	X	0	0	1	FNDC4
240139_at	0.019333767	0	X	X	0	2	SLC25A14
240180_at	1.06583E-05	0	X	0	X	2	ARIH1
240399_at	5.7363E-05	0	X	0	0	1	EMC4
240432_x_at	0.00041124	0	X	0	0	1	ST
240478_at	0.000120061	0	X	0	0	1	0
240830_at	0.005860075	0	X	0	X	2	0
240908_at	0.007039266	0	X	0	0	1	0
240921_at	0.007850817	0	X	0	0	1	GNPTAB
240971_x_at	0.000489823	0	X	0	0	1	SESTD1
241091_at	0.001453178	0	X	0	0	1	PARD3
241223_x_at	0.000715414	0	X	0	0	1	NUMA1
241303_x_at	0.000127111	0	X	0	0	1	TRIOBP
241505_at	0.008230906	0	X	0	0	1	COL22A1
241613_at	0.004407377	0	X	0	0	1	CO
241681_at	0.001164853	0	X	0	0	1	MAEA
241701_at	0.002935816	0	X	0	0	1	PPFIA1
241762_at	0.004798478	0	X	0	0	1	0
241769_at	0.005334516	0	X	0	X	2	NUCKS1
241797_at	0.000351192	0	X	0	0	1	0
241824_at	1.80696E-05	0	X	0	X	2	FRMD8
241851_x_at	0.00760227	0	X	0	0	1	KDSR
241997_at	0.009647229	0	X	0	X	2	DTNA
242014_at	0.000421681	0	X	0	0	1	ZBTB20
242068_at	0.002253767	0	X	0	0	1	LOC100130987
242131_at	0.01573594	0	X	0	0	1	RNF4
242235_x_at	0.000916907	0	X	0	0	1	TRIOBP
242265_at	0.001742709	0	X	0	0	1	PRSS3
242274_x_at	0.026253858	0	X	0	0	1	POLR3H
242287_at	0.000780256	0	X	0	0	1	CO
242303_at	3.60901E-05	0	X	0	0	1	PMS2P9
242326_at	0.020497489	0	0	0	X	1	AP2A2
242366_at	0.011930438	0	X	0	0	1	AKAP10
242372_s_at	0.006548976	0	X	0	0	1	0
242389_at	0.016563468	0	X	0	0	1	CBFA2T2
242428_at	0.000305716	0	X	0	0	1	GBP1
242431_at	0.002322299	0	X	0	0	1	NSA2
242467_at	0.001757031	0	X	X	0	2	0
242498_x_at	1.86868E-05	0	X	0	0	1	NCBP2-AS2
242558_at	4.92023E-05	0	X	0	0	1	IAH1
242608_x_at	0.000196333	0	X	0	0	1	0
242622_x_at	0.028996917	0	X	0	0	1	SLC25A42
242646_at	0.001186943	0	X	0	0	1	ID2
242742_at	0.002811325	0	X	0	0	1	CO
242748_at	3.33083E-05	0	X	0	0	1	TDRD3
242801_at	0.001036064	0	X	0	0	1	0

Sensitivity analysis result. (continued)

Probe ID	BF-value	CABk	EC	HIP	MTG-PC	DS	Symbol
242835_s_at	0.002945966	0	X	0	0	1	0
242872_at	6.31773E-05	0	0	X	0	1	SSU72
242916_at	0.001175741	0	X	0	0	1	0
242922_at	0.003599782	0	X	0	0	1	HMP19
242990_at	0.000934595	0	X	0	0	1	RBMS2
243001_at	0.003328953	0	X	0	0	1	PTPN13
243006_at	3.4377E-05	0	X	0	0	1	ADD3
243039_at	0.118461996	0	X	0	0	1	PTEN
243127_x_at	3.4577E-05	0	X	0	0	1	CLU
243158_at	3.4677E-05	0	X	0	0	1	0
243203_at	3.4777E-05	0	X	0	0	1	RAC1
243256_at	0.000298352	0	X	0	0	1	HTR2A
243295_at	3.4977E-05	0	X	0	0	1	USP54
243398_at	0.003182028	0	X	0	0	1	SMYD2
243496_at	3.5177E-05	0	X	0	0	1	0
243559_at	0.016801836	0	X	0	0	1	GNG2
243618_s_at	1.6175E-05	0	0	X	0	1	SRCIN1
243713_at	0.00014688	0	X	0	0	1	CDH12
243745_at	3.5477E-05	0	X	0	0	1	ITPRIPL2
243768_at	0.000246757	0	X	0	0	1	VPS35
243792_x_at	3.5677E-05	0	X	0	0	1	TBL1
243826_at	3.5777E-05	0	X	0	0	1	PARL
243878_at	3.5877E-05	0	X	0	0	1	IMP3
243988_at	0.002249758	0	X	0	0	1	ATP6V1C1
243993_at	3.6077E-05	0	X	0	0	1	SKIL
244040_at	0.002971084	0	X	0	0	1	COL5A3
244062_at	3.6277E-05	0	X	0	0	1	0
244128_x_at	0.000167201	0	X	X	0	2	ZNF131
244197_x_at	3.6477E-05	0	X	0	0	1	PTBP1
244292_at	0.000883044	0	X	0	0	1	0
244458_at	3.6677E-05	0	X	0	0	1	ATP1A3
244607_at	0.00562337	0	0	0	X	1	SFI1
244766_at	3.6777E-05	0	X	0	0	1	PTPRN2
31637_s_at	0.00361101	0	X	0	0	1	ITGB8
32099_at	3.6977E-05	0	X	0	0	1	IDH3B
37152_at	0.001220039	0	X	0	0	1	0
38069_at	5.77772E-05	0	0	X	X	2	CHI3L1
38398_at	3.7177E-05	0	X	0	X	2	LOC148413
38892_at	6.90774E-05	0	0	0	X	1	ZNF423
38964_r_at	0.003669736	0	X	0	0	1	0
40225_at	3.7377E-05	0	X	0	0	1	PPIA
41387_r_at	0.000579348	0	X	X	0	2	CEP41
47608_at	0.149904763	0	X	0	0	1	CRIM1
52005_at	0.000241644	0	X	0	0	1	KANK2
52255_s_at	0.051047711	0	X	0	0	1	PALD1
52731_at	0.00097584	0	X	0	0	1	SLC1A2
58780_s_at	0.052056615	0	X	0	X	2	TJAP1
58900_at	0.001373983	0	X	0	0	1	0
58994_at	6.33773E-05	0	0	X	X	2	PARD3B
65635_at	0.003168224	0	X	0	0	1	0

Probe ID is the list of probes. **BF-value** is the Bonferroni corrected value for each probe. **CABk** indicates, given as 'X', the probes that are in the Coloured (α, β)-k Feature Set problem approach result. **EC** indicates, given as 'X', the probes that are in the result of Coloured (α, β)-k Feature Set problem approach when the EC region is removed from the combined data. **HIP** indicates, given as 'X', the probes that are in the result of Coloured (α, β)-k Feature Set problem approach when the HIP region is removed from the combined data. **MTG-PC** indicates, given as 'X', the probes that are in the result of Coloured (α, β)-k Feature Set problem approach when the MTG and PC regions are removed from the combined data.**DS** indicates the the number of dataset. **Symbol** is the list of gene symbols.

Chapter 9

Appendix C

9.1 R source code used for the data analysis

R code used for the integration of the datasets produced using different platforms

This script addresses the aggregation of files for meta analysis, through the matching of different platform probes targetting the same gene. The matching uses the association of probe to gene to genomic position. For each gene, the probes targetting this gene are identified in each platform, and then combined in a "meta-feature". The goal is the generation of a combined file produced from the aggregation of individual experiments files.

```
Required Packages
> require(IRanges)
> require(cibm.utils)
> source("testAmbig.R")
> # source("writeABK.R")

        Importing the Datasets

> datasets <- list(
+ list(name="L2695",
+ rangeFile="problem-files/Lap-2695.ann-sorted",
+ abkFile="problem-files/2695-entF.abk"),
+ list(name="L3044",
+ rangeFile="problem-files/Lap-3044.ann-sorted",
+ abkFile="problem-files/3044-entF.abk"),
+ list(name="L3289",
+ rangeFile="problem-files/Lap-3289.ann-sorted",
+ abkFile="problem-files/3289-entF.abk"),
+ list(name="Welsh",
+ rangeFile="problem-files/Welsh.ann-sorted",
+ abkFile="problem-files/Welsh-entF.abk"),
+ list(name="Uma",
+ rangeFile="problem-files/Uma.ann-sorted",
+ abkFile="problem-files/Uma-entF.abk"),
+ list(name="Singh",
+ rangeFile="problem-files/Singh.ann-sorted",
+ abkFile="problem-files/Singh-entF.abk")
+ )
> outputFileName <- "combined-3"
```

```
> rangeData <- list()
> allRanges <- NULL
> totCols <- 0
> allColnames <- NULL
> allColclasses <- NULL
> allCaseColour <- NULL


Preprocessing of datasets


> curDS <- 1
> for (d in datasets) {
+ cat("Process",d[["name"]],":")
+ # Read the interval file for the dataset
+ rngd <- read.table(d[["rangeFile"]],header=F)
+ numRw <- dim(rngd)[1]
+ intv <- RangedData(IRanges(rngd[,5],rngd[,6]),
+ probeId=rngd[,1],
+ seqId=rngd[,2],
+ expId=d[["name"]],
+ mapIdx=0, # To fill later
+ space=rngd[,3]
+ )
+ ff <- file(d[["abkFile"]],open="r")
+ readLines(con=ff, n=3)
+ ll <- readLines(con=ff, n=1)
+ numRa <- as.integer(ll)
+ ll <- readLines(con=ff, n=1)
+ numC <- as.integer(ll)
+ ll <- readLines(con=ff, n=1)
+ nnam <- unlist(strsplit(ll,"\t"))
+ nnam <- nnam[1:numC+1] # Puzzling, why not [2:numC+1]
+ abkD <- read.table(ff,header=F,row.names=1,nrows=numRa)
+ colnames(abkD) <- nnam
+ ll <- readLines(con=ff, n=1)
+ ncla <- as.integer(unlist(strsplit(ll,"\t")))
+ ncla <- ncla[1:numC+1]
+ close(ff)
+
+ if (is.null(allRanges)) {
+ allRanges <- intv
+ allColnames <- paste(d[["name"]],":",nnam,sep="")
+ allColclasses <- ncla
+ allCaseColour <- rep(curDS,numC)
+ }
+ else {
+ allRanges <- rbind(allRanges,intv)
+ allColnames <- c(allColnames, paste(d[["name"]],":",nnam,sep=""))
+ allColclasses <- c(allColclasses, ncla)
+ allCaseColour <- c(allCaseColour, rep(curDS,numC))
+ }
+
+ totCols <- totCols + numC
+ cat(numRw,"intervals,",numRa,"features,",numC,
+ "samples.Full_list_size:",dim(allRanges)[1],"\n")
+
+ rangeData[[d[["name"]]]] <- list(
+ numRanges=numRw,
```

```
+ rngData=rngd,
+ rngList=intv,
+ numDataRows=numRa,
+ numDataCols=numC,
+ dataClass=ncla,
+ data=abkD
+ )
+
+ curDS <- curDS + 1

> geneRef <- read.table("refGene.txt", header=T, comment.char="")
> validIdx <- setdiff(1:length(geneRef$chrom), grep("_",as.vector(geneRef$chrom)))
> geneRef <- geneRef[validIdx,]
> mygrd <- RangedData(IRanges(geneRef$txStart,geneRef$txEnd),
+ space=geneRef$chrom,
+ symbol=geneRef$name2)

> myrgrd <- reduce(mygrd, min.gapwidth=1000L)
> ired <- as.matrix(findOverlaps(myrgrd, mygrd))
> iovr <- as.matrix(findOverlaps(myrgrd, allRanges))
> idxList <- unique(iovr[,1])

> auxSymbList <- as.vector(mygrd[["symbol"]])
> symbNames <- vector("character",length=length(idxList))
> for (k in 1:length(idxList)) {
+ idx2 <- idxList[k]
+ oriIdx <- ired[ired[,1] == idx2, 2]
+ if (length(oriIdx) > 0) {
+ nam <- c()
+ for (j in oriIdx) {
+ if (!(auxSymbList[j] %in% nam)) {
+ nam <- c(nam, auxSymbList[j])
+ }
+ }
+ symbNames[[k]] <- paste(nam,collapse=";")
+ } else {
+ symbNames[[k]] <- "error"
+ }
+ }

Getting the probes for the combined data

> masterRanges <- RangedData(IRanges(start(myrgrd)[idxList], end(myrgrd)[idxList]),
+ space=as.vector(space(myrgrd))[idxList],
+ symbol=symbNames)
> numMasterRanges <- dim(masterRanges)[1]
> cat("Number of master ranges:", numMasterRanges, "\n")


> masterRngProbeMap <- list()
> masterRngDataIndexCombs <- vector("list", length=numMasterRanges)
> for (d in datasets) {
+ cat("Process overlaps for: ",d[["name"]])
+ abkData <- rangeData[[d[["name"]]]][["data"]]
+ rns <- rownames(abkData)
+ rng <- rangeData[[d[["name"]]]]
+ iovr <- as.matrix(findOverlaps(rng[["rngList"]], masterRanges))
+ upIdx <- unique(iovr[,1])
+ ugIdx <- unique(iovr[,2])
```

```
+ cat("\tFrom ",rng[["numRanges"]],
+ "ranges,",length(upIdx),
+ "overlapover",length(ugIdx),"genes\n")
+  rangeData[[d[["name"]]]][["rngList"]][["mapIdx"]][iovr[,1]] <- iovr[,2]
+ maps <- list()
+ pids <- as.vector(rng[["rngList"]][["probeId"]])
+ for (k in 1:numMasterRanges) {
+ filt <- iovr[,2] == k
+ maps[[k]] <- unique(pids[iovr[filt ,1]])
+ nn <- length(maps[[k]])
+
+ if (nn > 0) {
+ myIdx <- match(maps[[k]], rns)
+ if (is.null(masterRngDataIndexCombs[[k]]))
+ masterRngDataIndexCombs[[k]] <- myIdx
+ else
+ masterRngDataIndexCombs[[k]] <- merge(masterRngDataIndexCombs[[k]], myIdx, + } else {
+ if (is.null(masterRngDataIndexCombs[[k]]))
+ masterRngDataIndexCombs[[k]] <- NA
+ else
+ masterRngDataIndexCombs[[k]] <- merge(masterRngDataIndexCombs[[k]], NA, by=NULL)
+ }
+ Debug prints
+  if (k %% 1000L == 0) {
+  cat(" R:",k,nn,"\n")
+  myIdx <- match(maps[[k]], rns)
+  print(myIdx)
+  print(masterRngDataIndexCombs[[k]])
+  }
+
+ }
+ masterRngProbeMap[[d[["name"]]]] <- maps
+ }

chk <- any(sapply(masterRngDataIndexCombs,function(x) all(is.na(x))))
> cat("++ Unmatched master ranges:",chk,"\n")
++ Unmatched master ranges: FALSE
> brow <- function(y) apply(y,1,function(x) !all(is.na(x)))
> chk <- any(sapply(masterRngDataIndexCombs,function(x) !all(brow(x))))
> cat("++ Any probe combination w/o indices:",chk,"\n")
++ Any probe combination w/o indices: FALSE
> masterRowCounts <- sapply(masterRngDataIndexCombs,function(x) sum(brow(x)))
> totRows <- sum(masterRowCounts)
> cat("Total number of master data rows: ",totRows,
+ " max:",max(masterRowCounts),
+ " min:",min(masterRowCounts),"\n")


        Data aggregation


> masterData <- vector("list",length=totRows)
> mDataIdx4mRngIdx <- cumsum(c(1,masterRowCounts))
> cat("[check--#ranges]",length(mDataIdx4mRngIdx)-1,"\n")
<datasetname>:<probeId>[|<datasetname>:<probeId>]
> masterRowNames <- sapply(1:totRows, function(x) paste("ci",x,sep=""))
> masterCompositeRowNames <- vector('character',length=totRows)
Do the loop:
> currCol <- 1
```

```
> currDs <- 1
> for (d in datasets) {
+ cat("Process data merge for: ",d[["name"]],":\n")
+ abkData <- rangeData[[d[["name"]]]][["data"]]
+ rns <- rownames(abkData)
+ myCols <- dim(abkData)[2]
+ nullRow <- rep(-1,myCols)

for (k in 1:numMasterRanges) {
+ # First and last row in data for this master range index
+ iniRw <- mDataIdx4mRngIdx[k]
+ endRw <- mDataIdx4mRngIdx[k+1] - 1
+
+ for (l in iniRw:endRw) {
+ # Index in data or NA
+ idx <- as.matrix(masterRngDataIndexCombs[[k]])[(l+1-iniRw),currDs]
+ if (is.na(idx)) {
+ masterData[[l]] <- c(masterData[[l]],nullRow)
+ }
+ else {
+ myRowData <- abkData[idx,1:myCols]
+ cat(sprintf(" Rng:%d\t%s\t@data row:%d : insert at [%d, %d:%d]\n",
+ k, rns[idx], idx, l, currCol, (currCol+myCols-1)))
+ masterData[[l]] <- c(masterData[[l]], myRowData)
+ masterCompositeRowNames[l] <- paste(
+ masterCompositeRowNames[l],
+ "|",d[["name"]],":",rns[idx],sep="")
+
+ }
+ }
+ }
+ currCol <- currCol + myCols
+ currDs <- currDs + 1
+ }

> masterGeneRowNames <- vector('character',length=totRows)
> for (k in 1:numMasterRanges) {
+ iniRw <- mDataIdx4mRngIdx[k]
+ endRw <- mDataIdx4mRngIdx[k+1] - 1
+
+ for (l in iniRw:endRw) {
+ masterGeneRowNames[l] <- symbNames[k]
+ }
+ }
> finalData <- matrix(-1,nrow=totRows,ncol=totCols)
> npc <- totRows %/% 10
> for (k in 1:totRows) {
+ finalData[k,] <- as.vector(unlist(masterData[[k]]))
+ if (k %% npc == 0) cat(k/npc * 10,"%\n")
+ }


> ambig <- testAmbig(finalData, allColclasses, allColnames, na.val=-1,\\ print.messages=F
> if (!is.null(ambig)) {
+
+ cat("Ambiguous columns !!\n")
+
+ } else {
+
```

```
+ aggData <- new("cibm.abk")
+ aggData@data <- as.data.frame(finalData)
+ rownames(aggData@data) <- paste("sr",1:totRows,sep="")
+ names(aggData@data) <- allColnames
+ labels(aggData) <- allColClasses
+ aggData@caseAttr <- data.frame(casecolours=allCaseColour,
+ caseweights=rep(1,totCols))
+ aggData@featureAttr <- data.frame(featureweights=rep(1,totRows))
+ ofn <- paste(outputFileName,".abk",sep="")
+ write.abk(aggData,file=ofn,gzip=F,out.equalweights=T)
+
+ fo <- file(paste(outputFileName,"-expanded-rows.txt",sep=""),open="w")
+ writeLines(paste(paste("sr", 1:totRows, sep=""),
+ masterCompositeRowNames,
+ masterGeneRowNames,sep="\t"),
+ con=fo)
+ close(fo)
+
+ }
```

Writing the combined data and realting files

```
> filt <- rep(TRUE,totCols)
> filt[unique(ambig[,2])] <- FALSE
> ambig <- testAmbig(finalData[,filt],allColclasses[filt],
+ allColnames[filt],na.val=-1)
> stopifnot(is.null(ambig))

> aggData <- new("cibm.abk")
> aggData@data <- as.data.frame(finalData[,filt])
> rownames(aggData@data) <- paste("sr",1:totRows,sep="")
> names(aggData@data) <- allColnames[filt]
> labels(aggData) <- allColclasses[filt]
> aggData@caseAttr <- data.frame(casecolours=allCaseColour[filt],
+ caseweights=rep(1,sum(filt)))
> aggData@featureAttr <- data.frame(featureweights=rep(1,totRows))
> ofn <- paste(outputFileName,".abk",sep="")
> write.abk(aggData,file=ofn,gzip=F,out.equalweights=T)
> fo <- file(paste(outputFileName,"-expanded-rows.txt",sep=""),open="w")
> writeLines(paste(paste("sr", 1:totRows, sep=""),
+ masterCompositeRowNames,
+ masterGeneRowNames,sep="\t"),
+ con=fo)
> close(fo)
```

The code used for the annotation of the list of probe IDs using the online biomaRt database against ENSEMBL and also looking for non-coding RNAs.

```
library(biomaRt)
library(xtable)
options(xtable.comment = FALSE)
options(xtable.booktabs = TRUE)
plist <- read.table(file='../Data//six-regions.out.sol', blank.lines.skip=T,
header=T, stringsAsFactors=F, sep='\t')
colnames(plist) <- c('features')
```

```
plist$features <- sub("^X", "", plist$features)
mart <- useMart("ensembl")
mart <- useDataset("hsapiens_gene_ensembl", mart=mart)
filter <- c('affy_hg_u133_plus_2')
annotation <- data.frame(probes=plist$features, stringsAsFactors=F)
myAttr1 <- c('affy_hg_u133_plus_2',
'ensembl_gene_id', 'description',
'external_gene_name', 'external_gene_source')
annot1 <- getBM(myAttr1, filters=filter, values=plist$features, mart)
annotation <- tablify(annotation, annot1, 1, 1)
myAttr2 <- c('affy_hg_u133_plus_2', 'mirbase_accession', 'mirbase_id')
annot2 <- getBM(myAttr2, filters=filter, values=plist$features, mart)
annotation <- tablify(annotation, annot2, 1, 1)
myAttr3 <- c('affy_hg_u133_plus_2', 'refseq_ncrna', 'refseq_ncrna_predicted')
annot3 <- getBM(myAttr3, filters=filter, values=plist$features, mart)
annotation <- tablify(annotation, annot3, 1, 1)
myAttr4 <- c('affy_hg_u133_plus_2', 'entrezgene', 'entrezgene_transcript_name')
annot4 <- getBM(myAttr4, filters=filter, values=plist$features, mart)
annotation <- tablify(annotation, annot4, 1, 1)
```

R code used for the pre-processing of AD datasets

```
regns <- list(list(pfx='EC', data=data.frame(), stats=data.frame()),\\
list(pfx='HIP', data=data.frame(), stats=data.frame()),\\
list(pfx='MTG', data=data.frame(), stats=data.frame()),\\
list(pfx='PC', data=data.frame(), stats=data.frame()),\\
list(pfx='SFG', data=data.frame(), stats=data.frame()),\\
list(pfx='VCX', data=data.frame(), stats=data.frame()))

for(p in 1:length(regns)) {
fnames <- list.files(path='../IndividualSamples/',
pattern=paste(regns[[p]]$pfx, '*', sep=' '))
sns <- c()
sts <- c()
cls <- c()
stats <- NULL
cat('Region:', regns[[p]]$pfx, length(fnames), 'files\n')
for(fname in fnames) {
dd <- read.table(file=paste('../IndividualSamples/', fname, sep=''),
header=F, skip=5, row.names=1,
col.names=c('probeId', 'value', 'type', 'pval'),
stringsAsFactors=F, flush=T)
sn <- make.names(sub('.txt', '', fname, fixed=T))
sns <- c(sns, sn)
sts <- c(sts, paste('T.', sn, sep=''), paste('pv.', sn, sep=''))
if(grepl(" AD ", fname))

cls <- c(cls, "AD")
else
cls <- c(cls, "Control")
regns[[p]]$data <- rbind(regns[[p]]$data, dd$value)
if(is.null(stats))
stats <- cbind(dd$type, dd$pval)
else
stats <- cbind(stats, dd$type, ddpval)
colnames(stats) <- sts
rownames(stats) <- rownames(dd)
regns[[p]]$stats <- stats
```

```
rownames(regns[[p]]$data) <- sns
colnames(regns[[p]]$data) <- rownames(dd)
# Remove control probes
rcp <-grep('^AFFX', rownames(dd), value=F)
regns[[p]]$data <- cbind(regns[[p]]data[,rcp], class=as.factor(cls))
}

 Entropy filtering

require(cibm.utils)
args\_bf <- c("-jar",
path.expand("~/tools/basicFilter/basicFilter-3.2rc3.jar"),
"-i", "(none)",
"-p", "entf", "-xm", "-xa",
"-k", "(none)")
abk\_data <- NULL
for(p in 1:length(regns)) {

nbifile <- paste('../Data/', regns[[p]]pfx, '\_ad.nbi', sep='')
abkfile <- paste('../Data/', regns[[p]]pfx, '\_ad.abk', sep='')

dmd <- dim(regns[[p]]data)
ddm <- t(regns[[p]]data[,1:(dmd[2]-1)])
ddm.sum <- colSums(ddm, na.rm=T)
ddm <- sweep(ddm, 2, ddm.sum, "/")
ddm <- rbind(ddm, regns[[p]]data[,dmd[2]])

 Add class and write nbi files

rownames(ddm)[dmd[2]] <- "class"
write.nbi(ddm, nbifile, featuresInRows=T)

Entropy filter and abk file

args\_bf[4] <- paste(nbifile, '.gz', sep='')
args\_bf[10] <- abkfile
tt <- system.time(
system2("java", args=args\_bf,
stdout=paste('../Data/', regns[[p]]pfx, '\_entf.log', sep=''))
)
cat(regns[[p]]pfx, ' Entropy Time:', tt)

abkd <- read.abk(file=abkfile)
abk\_data <- c(abk\_data,
list(list(name=regns[[p]]pfx, abkd=abkd, samps=dmd[1])))
}

Combined Analysis

probelist <- c()
ens <- c('probes')
for(d in abk_data) {
probelist <- unique(sort(c(probelist, rownames(d$abkd$data))))
}
probelist <- data.frame(probes=probelist, stringsAsFactors = F)
for(d in abk_data) {
probelist <- cbind(probelist, probelist$probes %in% rownames(d$abkd$data))
ens <- c(ens, d$name)
}
```

```
colnames(probelist) <- ens
cover <- apply(probelist[,2:length(ens)], 1, sum)
probelist <- cbind(probelist, cover=cover)

minCover <- 4
rns <- probelist$probes[which(probelist$cover >= minCover)]
combDS <- data.frame(row.names=rns, stringsAsFactors=F)
caseColours <- c()
caseClasses <- c()
for(p in abk_data) {
rwi <- match(rns, rownames(p$abkd$data))
combDS <- cbind(combDS, p$abkd$data[rwi, ])
caseColours <- c(caseColours, rep(p$name, times=length(p$abkd$classes)))
caseClasses <- c(caseClasses, p$abkd$classes)
}
combDS[is.na(combDS)] <- -1
finalABK <- get.default.abk(data=combDS,
classes=as.factor(caseClasses),
casecolours=caseColours,
alfaadjacency='true',
betaadjacency='true')
write.abk2(finalABK, file='../Data/six-regions.cabk', gzip=F)

Apply \CABkFS problem approach

args_abk <- c("-i", "../Data/six-regions.cabk", "-o", "../Data/six-regions.out",
"-m", "2", "-O", "max:feature_degree", "-y", "alfa+beta",
"--memoryEmphasis", "--useExternalNodes", "--saveIncumbents",
"--workingMemory", "256")
tt <- system.time(
system2(path.expand("~/tools/FABCPP/fabcpp-1.7.1"),
args=args_abk,
stdout='../Data/six-regions.abk.log',
stderr='../Data/six-regions.abk.err'
)
cat('\n', args_abk[2], " CABk Time:", tt, "\n")
```

Bibliography

[1] Heather A Adams et al. "Meta-analysis of genome-wide expression patterns associated with behavioral maturation in honey bees". In: *BMC Genomics* 9 (Oct. 2008), pp. 503–503. ISSN: 1471-2164. DOI: 10.1186/1471-2164-9-503.

[2] Alex A. Adjei. "Blocking Oncogenic Ras Signaling for Cancer Therapy". In: *Journal of the National Cancer Institute* 93.14 (2001), pp. 1062–1074. eprint: http://jnci.oxfordjournals.org/content/93/14/1062.full.pdf+html.

[3] Smriti Agarwal, Rudi K Tannenberg and Peter R Dodd. "Reduced expression of the inhibitory synapse scaffolding protein gephyrin in Alzheimer's disease". In: *Journal of Alzheimer's Disease* 14.3 (2008), pp. 313–322.

[4] FΓʀtima Al-Shahrour et al. "FatiGO +: a functional profiling tool for genomic data. Integration of functional annotation, regulatory motifs and interaction data with microarray experiments". In: *Nucleic Acids Research* 35.suppl 2 (2007), W91–W96. DOI: 10.1093/nar/gkm260. eprint: http://nar.oxfordjournals.org/content/35/suppl_2/W91.full.pdf+html.

[5] Ben-Dor Amir et al. "Tissue Classification with Gene Expression Profiles." In: *Journal of Computational Biology.* 7(3-4) (2004), pp. 559 –583. DOI: 10.1089/106652700750050943..

[6] Vijayalakshmi Ananthanarayanan et al. "Alpha-methylacyl-CoA racemase (AMACR) expression in normal prostatic glands and high-grade prostatic intraepithelial neoplasia (HGPIN): Association with diagnosis of prostate cancer". In: *The Prostate* 63.4 (2005), pp. 341–346. ISSN: 1097-0045. DOI: 10.1002/pros.20196.

[7] Caitlin Andrews and Peter A. Humphrey. "Utility of ERG Versus AMACR Expression in Diagnosis of Minimal Adenocarcinoma of the Prostate in Needle Biopsy Tissue". In: *The American Journal of Surgical Pathology* 38.7 (2014), pp. 1007–1012. ISSN: 0147-5185. DOI: 10.1097/PAS.0000000000000205.

[8] Daphne Antoniou, Athanasios Stergiopoulos and Panagiotis K Politis. "Recent advances in the involvement of long non-coding RNAs in neural stem cell biology and brain pathophysiology". In: *Frontiers in Physiology* 5 (Apr. 2014), pp. 155–. ISSN: 1664-042X.

[9] Ahmed Shamsul Arefin et al. "Unveiling Clusters of RNA Transcript Pairs Associated with Markers of Alzheimer's Disease Progression". In: *PLoS ONE* 7.9 (Sept. 2012), e45535. DOI: 10.1371/journal.pone.0045535.

[10] Soline Aubry et al. "Assembly and Interrogation of AlzheimerвЂ™s Disease Genetic Networks Reveal Novel Regulators of Progression". In: *PLoS ONE* 10.3 (Jan. 2015), e0120352–. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0120352.

[11] Ingrid Van der Auwera et al. "A ketogenic diet reduces amyloid beta 40 and 42 in a mouse model of Alzheimer's disease". In: *Nutrition & Metabolism* 2.1 (2005), p. 28. ISSN: 1743-7075. DOI: 10.1186/1743-7075-2-28.

[12] John C. Bailar. "The Promise and Problems of Meta-Analysis". In: *New England Journal of Medicine* 337.8 (1997). PMID: 9262502, pp. 559–561. DOI: 10.1056/NEJM199708213370810. eprint: http://dx.doi.org/10.1056/NEJM199708213370810.

[13] Pierre Baldi and Anthony D. Long. "A Bayesian framework for the analysis of microarray expression data: regularized t -test and statistical inferences of gene changes". In: *Bioinformatics* 17.6 (2001), pp. 509–519. DOI: 10.1093/bioinformatics/17.6.509. eprint: http://bioinformatics.oxfordjournals.org/content/17/6/509.full.pdf+html.

[14] Tanya Barrett et al. "NCBI GEO: archive for functional genomics data sets-update". In: *Nucleic Acids Research* 41.D1 (2013), pp. D991–D995. DOI: 10.1093/nar/gks1193. eprint: http://nar.oxfordjournals.org/content/41/D1/D991.full.pdf+html.

[15] Gary W. Beecham et al. "Genome-Wide Association Meta-analysis of Neuropathologic Features of Alzheimer's Disease and Related Dementias". In: *PLoS Genet* 10.9 (Sept. 2014), e1004606. DOI: 10.1371/journal.pgen.1004606.

[16] Goran Benko et al. "Impact of the EpCAM expression on biochemical recurrence free survival in clinically localized prostate cancer". In: *Urologic Oncology: Seminars and Original Investigations* 31.4 (2013), pp. 468 –474. ISSN: 1078-1439. DOI: http://dx.doi.org/10.1016/j.urolonc.2011.03.007.

[17] David R Bentley et al. "Accurate Whole Human Genome Sequencing using Reversible Terminator Chemistry". In: *Nature* 456.7218 (Nov. 2008), pp. 53–59. ISSN: 1476-4687. DOI: 10.1038/nature07517.

[18] Dominic B Bernkopf and Elizabeth D Williams. "Potential role of EPB41L3 (Protein 4.1B/Dal-1) as a target for treatment of advanced prostate cancer". In: *Expert Opinion on Therapeutic Targets* 12.7 (2008), pp. 845–853. DOI: doi:10.1517/14728222.12.7.845.

[19] R. Berretta and P. Moscato. "Cancer biomarker discovery: the entropic hallmark". In: *PloS one* 5.8 (2010), e12262. ISSN: 1932-6203 (Electronic) 1932-6203 (Linking). DOI: 10.1371/journal.pone.0012262.

[20] Regina Berretta, Wagner Costa and Pablo Moscato. "Combinatorial Optimization Models for Finding Genetic Signatures from Gene Expression Datasets". In: *Bioinformatics*. Ed. by JonathanM Keith. Vol. 453. Humana Press, 2008, pp. 363–377. ISBN: 978-1-60327-428-9.

[21] Regina Berretta, Alexandre Mendes and Pablo Mascato. "Selection of Discriminative Genes in Microarray Experiments Using Mathematical Programming." In: *Journal of Research & Practice in Information Technology* 39.4 (2007). ABk, pp. 287 –299. ISSN: 1443458X.

[22] Lars Bertram et al. "Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database". In: *Nat Genet* 39.1 (Jan. 2007), pp. 17–23. ISSN: 1061-4036. DOI: 10.1038/ng1934.

[23] M. Bigl et al. "Activities of key glycolytic enzymes in the brains of patients with Alzheimer's disease". English. In: *Journal of Neural Transmission* 106.5-6 (1999), pp. 499–511. ISSN: 0300-9564. DOI: 10.1007/s007020050174.

[24] E.M. Blalock et al. "Harnessing the power of gene microarrays for the study of brain aging and Alzheimer's disease: Statistical reliability and functional correlation". In: *Ageing Research Reviews* 4.4 (2005). Gene Expression During Aging, pp. 481 –512. ISSN: 1568-1637. DOI: http://dx.doi.org/10.1016/j.arr.2005.06.006.

[25] Kaj Blennow, Mony J de Leon and Henrik Zetterberg. "Alzheimer's disease". In: *The Lancet* 368.9533 (2006), pp. 387 –403. ISSN: 0140-6736. DOI: http://dx.doi.org/10.1016/S0140-6736(06)69113-7.

[26] Trond Bo and Inge Jonassen. "New feature subset selection procedures for classification of expression profiles". In: *Genome biology* 3.4 (2002), pp. 1–0017.

[27] Foucaud du Boisgueheneuc et al. "Functions of the left superior frontal gyrus in humans: a lesion study". In: *Brain* 129.12 (2006), pp. 3315–3328. ISSN: 0006-8950. DOI: 10.1093/brain/awl244.

[28] H. Braak and E. Braak. "Neuropathological stageing of Alzheimer-related changes". English. In: *Acta Neuropathologica* 82.4 (1991), pp. 239–259. ISSN: 0001-6322. DOI: 10.1007/BF00308809.

[29] R. Bradbury and M.A. Brodney. *Alzheimer's Disease*. Topics in Medicinal Chemistry. Springer Berlin Heidelberg, 2007, p. 179. ISBN: 9783540742296.

[30] Rainer Breitling et al. "Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments". In: *FEBS Letters* 573.1–3 (2004), pp. 83 –92. ISSN: 0014-5793. DOI: http://dx.doi.org/10.1016/j.febslet.2004.07.055.

[31] Eva M F Brekke et al. "Quantitative importance of the pentose phosphate pathway determined by incorporation of 13C from [2-13C]- and [3-13C] glucose into TCA cycle intermediates and neurotransmitter amino acids in functionally intact neurons". In: *J Cereb Blood Flow Metab* 32.9 (Sept. 2012), pp. 1788–1799. ISSN: 0271-678X. DOI: 10.1038/jcbfm.2012.85.

[32] Alyssa A Brewer and Brian Barton. "Visual cortex in aging and Alzheimer's disease: changes in visual field maps and population receptive fields". In: *Frontiers in Psychology* 5 (Jan. 2014), pp. 74–. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2014.00074.

[33] Holly Bridge et al. "Changes in connectivity after visual cortical brain damage underlie altered visual function". In: *Brain* 131.6 (2008), pp. 1433–1444. ISSN: 0006-8950. DOI: 10.1093/brain/awn063.

[34] Wendy M. Brooks et al. "Gene expression profiles of metabolic enzyme transcripts in Alzheimer's disease". In: *Brain Research* 1127 (Jan. 2007), pp. 127–135. ISSN: 0006-8993. DOI: 10.1016/j.brainres.2006.09.106.

[35] Andreas Buness et al. "arrayMagic: two-colour cDNA microarray quality control and preprocessing". In: *Bioinformatics* 21.4 (2005), pp. 554–556. DOI: 10.1093/bioinformatics/bti052. eprint: http://bioinformatics.oxfordjournals.org/content/21/4/554.full.pdf+html.

[36] N. Burns-Cox et al. "Changes in collagen metabolism in prostate cancer: a host response that may alter progression". In: *The Journal of Urology* 166.5 (2001), pp. 1698–1701. ISSN: 0022-5347. DOI: http://dx.doi.org/10.1016/S0022-5347(05)65656-X.

[37] S. Caggia et al. "Modulation of YY1 and p53 expression by transforming growth factor-Oÿ3 in prostate cell lines". In: *Cytokine* 56 (2011), 403вГУ410.

[38] Patrick Cahan et al. "Meta-analysis of microarray results: challenges, opportunities, and recommendations for standardization". In: *Gene* 401 (2007), pp. 12 –18. ISSN: 0378-1119. DOI: 10.1016/j.gene.2007.06.016.

[39] Frederic Checler et al. "Role of the proteasome in Alzheimer's disease". In: *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1502.1 (July 2000), pp. 133–138. ISSN: 0925-4439. DOI: 10.1016/S0925-4439(00)00039-9.

[40] Ling-Ling Chen and Gordon G Carmichael. "Decoding the function of nuclear long noncoding RNAs". In: *Current opinion in cell biology* 22.3 (Mar. 2010), pp. 357–364. ISSN: 1879-0410.

[41] Jung Kyoon Choi et al. "Combining multiple microarray studies and modeling interstudy variation". In: *Bioinformatics* 19.suppl 1 (2003), pp. i84–i90. DOI: 10.1093/bioinformatics/btg1010. eprint: http://bioinformatics.oxfordjournals.org/content/19/suppl_1/i84.full.pdf+html.

[42] Jung Kyoon Choi et al. "Integrative analysis of multiple gene expression profiles applied to liver cancer study". In: *FEBS Letters* 565.1ГђBЇBY3 (2004), pp. 93 –100. ISSN: 0014-5793. DOI: http://dx.doi.org/10.1016/j.febslet.2004.03.081.

[43] L.W.K. Chung, W.B. Isaacs and J.W. Simons. *Prostate Cancer: Biology, Genetics, and the New Therapeutics*. Contemporary Cancer Research. Humana Press, 2007. ISBN: 9781597452243.

[44] A Convit et al. "Atrophy of the medial occipitotemporal, inferior, and middle temporal gyri in non-demented elderly predict decline to Alzheimer's disease". In: *Neurobiology of Aging* 21.1 (2000), pp. 19 –26. ISSN: 0197-4580. DOI: http://dx.doi.org/10.1016/S0197-4580(99)00107-4.

[45] Carlos Cotta and Pablo Moscato. "The k-Feature Set problem is W[2]-complete". In: *Journal of Computer and System Sciences* 67.4 (2003). Parameterized Computation and Complexity 2003, pp. 686 –690. ISSN: 0022-0000. DOI: 10.1016/S0022-0000(03)00081-3.

[46] Carlos Cotta, Christian Sloper and Pablo Moscato. "Evolutionary Search of Thresholds for Robust Feature Set Selection: Application to the Analysis of Microarray Data". In: *Applications of Evolutionary Computing*. Ed. by Ganther Raidl et al. Vol. 3005. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2004, pp. 21–30. ISBN: 978-3-540-21378-9.

[47] Luisa Cutillo, Annamaria Carissimo and Silvia Figini. "Network Selection: A Method for Ranked Lists Selection". In: *PLoS ONE* 7.8 (Aug. 2012), e43678. DOI: 10.1371/journal.pone.0043678.

[48] Roy G Cutler et al. "Involvement of oxidative stress-induced abnormalities in ceramide and cholesterol metabolism in brain aging and Alzheimer's disease". In: *Proceedings of the National Academy of Sciences of the United States of America* 101.7 (Sept. 2003), pp. 2070–2075. ISSN: 1091-6490. DOI: 10.1073/pnas.0305799101.

[49] Scott Davies and Stuart Russell. "NP-completeness of searches for smallest possible feature sets". In: *Proceedings of the 1994 AAAI fall symposium on relevance*. Proceedings of the 1994 AAAI fall symposium on relevance. AAAI Press. 1994, pp. 37–39.

[50] Robert P DeConde et al. "Combining Results of Microarray Experiments: A Rank Aggregation Approach". In: *Statistical Applications in Genetics and Molecular Biology* 5 (2006), pp. 1544–6115. DOI: 10.2202/1544-6115.1204.

[51] X. Delbeuck, M. Van der Linden and F. Collette. "Alzheimer' Disease as a Disconnection Syndrome?" English. In: *Neuropsychology Review* 13.2 (2003), pp. 79–92. ISSN: 1040-7308. DOI: 10.1023/A:1023832305702.

[52] Ramon Diaz-Uriarte and Sara Alvarez de Andres. "Gene selection and classification of microarray data using random forest". In: *BMC Bioinformatics* 7.1 (2006), p. 3. ISSN: 1471-2105. DOI: 10.1186/1471-2105-7-3.

[53] Bingqian Ding et al. "Gene Expression Profiles of Entorhinal Cortex in Alzheimer's Disease". In: *American Journal of Alzheimer's Disease and Other Dementias* 29.6 (2014), pp. 526–532. DOI: 10.1177/1533317514523487. eprint: http://aja.sagepub.com/content/29/6/526.full.pdf+html.

[54] Chris H.Q. Ding. "Unsupervised feature selection via two-way ordering in gene expression analysis". In: *Bioinformatics* 19.10 (2003), pp. 1259–1266. DOI: 10.1093/bioinformatics/btg149. eprint: http://bioinformatics.oxfordjournals.org/content/19/10/1259.full.pdf+html.

[55] Chris Ding and Hanchuan Peng. "Minimum redundancy feature selection from microarray gene expression data." In: *Journal of Bioinformatics and Computational Biology* 03.02 (2005), pp. 185–205. DOI: 10.1142/S0219720005001004. eprint: http://www.worldscientific.com/doi/pdf/10.1142/S0219720005001004.

[56] Yan Dong et al. "Androgen receptor signaling intensity is a key factor in determining the sensitivity of prostate cancer cells to selenium inhibition of growth and cancer–specific biomarkers". In: *Molecular Cancer Therapeutics* 4.7 (2005), pp. 1047–1055. DOI: 10.1158/1535-7163.MCT-05-0124. eprint: http://mct.aacrjournals.org/content/4/7/1047.full.pdf+html.

[57] R. Rod G. Downey and M.R. Fellows. *Parameterized Complexity*. Monographs in Computer Science. Springer Verlag, 1999. ISBN: 9780387948836.

[58] Epaminondas Doxakis. "Post-transcriptional Regulation of Oç-Synuclein Expression by mir-7 and mir-153". In: *The Journal of Biological Chemistry* 285.17 (Jan. 2010), pp. 12726–12734. ISSN: 1083-351X. DOI: 10.1074/jbc.M109.086827.

[59] Michal Draminski et al. "Monte Carlo feature selection for supervised classification". In: *Bioinformatics* 24.1 (2008), pp. 110–117. DOI: 10.1093/bioinformatics/btm486. eprint: http://bioinformatics.oxfordjournals.org/content/24/1/110.full.pdf+html.

[60] D; Draper et al. *Combining Information: Statistical Issues and Opportunities for Research*. Contemporary Statistics. National Academy Press, 1992. ISBN: 9780309047302.

[61] Pan Du, Warren A. Kibbe and Simon M. Lin. "lumi: a pipeline for processing Illumina microarray". In: *Bioinformatics* 24.13 (2008), pp. 1547–1548. DOI: 10.1093/bioinformatics/btn224. eprint: http://bioinformatics.oxfordjournals.org/content/24/13/1547.full.pdf+html.

[62] Luc Duchateau et al. "Individual Patient–versus Literature-Based Meta-analysis of Survival Data: Time to Event and Event Rate at a Particular Time Can Make a Difference, an Example Based on Head and Neck Cancer". In: *Controlled Clinical Trials* 22.5 (2001), pp. 538 –547. ISSN: 0197-2456. DOI: 10.1016/S0197-2456(01)00152-0.

[63] Sandrine Dudoit, Juliet Popper Shaffer and Jennifer C. Boldrick. "Multiple Hypothesis Testing in Microarray Experiments". In: *Statistical Science* 18(1) (2003), pp. 71–103. ISSN: 0883-4237. DOI: 10.1214/ss/1056397487.

[64] BΓ ̨atrice Duval and Jin-Kao Hao. "Advances in metaheuristics for gene selection and classification of microarray data". In: *Briefings in Bioinformatics* 11.1 (2010), pp. 127–141. DOI: 10.1093/bib/bbp035. eprint: http://bib.oxfordjournals.org/content/11/1/127.full.pdf+html.

[65] Darius M. Dziuda. "Basic Analysis of Gene Expression Microarray Data". In: *Data Mining for Genomics and Proteomics*. John Wiley & Sons, Inc., 2010, pp. 17–93. ISBN: 9780470593417. DOI: 10.1002/9780470593417.ch2.

[66] Darius M. Dziuda. "Biomarker Discovery and Classification". In: *Data Mining for Genomics and Proteomics*. John Wiley & Sons, Inc., 2010, pp. 95–200. ISBN: 9780470593417. DOI: 10.1002/9780470593417.ch3.

[67] Science Made Easy. *Microbes in prostate cancer and Kawasaki disease*. 2015. URL: http://sciencemadeeasy.kinja.com/microbes-in-prostate-cancer-and-kawasaki-disease-1578821419.

[68] Matthias Egger and George Davey Smith. "Meta-analysis: Potentials and promise". In: *BMJ* 315.7119 (1997), pp. 1371–1374. ISSN: 0959-8138. DOI: 10.1136/bmj.315.7119.1371.

[69] Liat Ein-Dor et al. "Outcome signature genes in breast cancer: is there a unique set?" In: *Bioinformatics* 21.2 (2005), pp. 171–178. DOI: 10.1093/bioinformatics/bth469. eprint: http://bioinformatics.oxfordjournals.org/content/21/2/171.full.pdf+html.

[70] Ashraf El Fiky, Allison E. Arch and John J. Krolewski. "Intracellular domain of the IFNaR2 interferon receptor subunit mediates transcription via Stat2". In: *Journal of Cellular Physiology* 204.2 (2005), pp. 567–573. ISSN: 1097-4652. DOI: 10.1002/jcp.20305.

[71] William J Elliott and Peter M Meyer. "Incident diabetes in clinical trials of antihypertensive drugs: a network meta-analysis". In: *The Lancet* 369.9557 (2007), pp. 201 –207. ISSN: 0140-6736. DOI: http://dx.doi.org/10.1016/S0140-6736(07)60108-1.

[72] Tobias Engl et al. "Prostate tumor CXC-chemokine profile correlates with cell adhesion to endothelium and extracellular matrix". In: *Life Sciences* 78.16 (2006), pp. 1784–1793. ISSN: 0024-3205. DOI: http://dx.doi.org/10.1016/j.lfs.2005.08.019.

[73] S Hossein Fatemi et al. "Expression of GABA(B) receptors is altered in brains of subjects with autism". In: *Cerebellum (London, England)* 8.1 (Mar. 2009), pp. 64–69. ISSN: 1473-4230. DOI: 10.1007/s12311-008-0075-3.

[74] M. Faust. *The Handbook of the Neuropsychology of Language*. Ed. by Miriam Faust. Vol. Volume 2 of Blackwell Handbooks of Behavioral Neuroscience. Blackwell Handbooks of Behavioral Neuroscience. Wiley, 2015, p. 1056. ISBN: 9781119050469. DOI: ISBN1119050464,9781119050469.

[75] Usama M. Fayyad and Keki B. Irani. "Multi-Interval Discretization of Continuous-Valued Attributes for Classification Learning". In: *Proceedings of the 13th International Joint Conference on Artificial Intelligence. Chambéry, France, August 28 - September 3, 1993*. 1993, pp. 1022–1029.

[76] F Fernandez-Enright et al. "Novel implications of Lingo-1 and its signaling partners in schizophrenia". In: *Transl Psychiatry* 4 (Jan. 2014), e348. DOI: 10.1038/tp.2013.121.

[77] A. Fernandez-Medarde et al. "Laser microdissection and microarray analysis of the hippocampus of Ras-GRF1 knockout mice reveals gene expression changes affecting signal transduction pathways related to memory and learning". In: *Neuroscience* 146.1 (Apr. 2007), pp. 272–285. ISSN: 0306-4522.

[78] Massimo S. Fiandaca et al. "The critical need for defining preclinical biomarkers in Alzheimer's disease". In: *Alzheimer's & Dementia* 10.3, Supplement (2014). Military Risk Supplement, S196 –S212. ISSN: 1552-5260. DOI: http://dx.doi.org/10.1016/j.jalz.2014.04.015.

[79] Sarah K. Fineberg, Kenneth S. Kosik and Beverly L. Davidson. "MicroRNAs Potentiate Neural Development". In: *Neuron* 64.3 (2009), pp. 303–309. ISSN: 0896-6273. DOI: http://dx.doi.org/10.1016/j.neuron.2009.10.020.

[80] Irit Fishel, Alon Kaufman and Eytan Ruppin. "Meta-analysis of gene expression data: a predictor-based approach". In: *Bioinformatics* 23.13 (2007), pp. 1599–1606. DOI: 10.1093/bioinformatics/btm149. eprint: http://bioinformatics.oxfordjournals.org/content/23/13/1599.full.pdf+html.

[81] Ronald Aylmer Sir Fisher. *Statistical methods for research workers*. 7th ed., rev. and enl. Edinburgh Oliver and Boyd, 1938, p. 356.

[82] Annette L Fitzpatrick et al. "Mid- and Late-Life Obesity: Risk of Dementia in the Cardiovascular Health Cognition Study". In: *Archives of neurology* 66.3 (Mar. 2009), pp. 336–342. ISSN: 1538-3687. DOI: 10.1001/archneurol.2008.582.

[83] J. Flum. *Parameterized Complexity Theory*. Texts in Theoretical Computer Science. An EATCS Series. Springer, 2006. ISBN: 9783540299530.

[84] D. Fong et al. "Expression of EpCAM(MF) and EpCAM(MT) variants in human carcinomas". In: *Journal of clinical pathology* 67.5 (2014), pp. 408–414. ISSN: 1472-4146 (Electronic) 0021-9746 (Linking). DOI: 10.1136/jclinpath-2013-201932.

[85] Richard Fox and Matthew Dimmic. "A two-sample Bayesian t-test for microarray data". In: *BMC Bioinformatics* 7.1 (2006), p. 126. ISSN: 1471-2105. DOI: 10.1186/1471-2105-7-126.

[86] Apostolia Fragkouli and Epaminondas Doxakis. "miR-7 and miR-153 protect neurons against MPP^+-induced cell death via upregulation of mTOR pathway". In: *Frontiers in Cellular Neuroscience* 8.182 (2014). ISSN: 1662-5102. DOI: 10.3389/fncel.2014.00182.

[87] Erik Fransen. "Functional role of entorhinal cortex in working memory processing". In: *Neural Networks* 18.9 (2005). Computational Theories of the Functions of the Hippocampus, pp. 1141 –1149. ISSN: 0893-6080. DOI: http://dx.doi.org/10.1016/j.neunet.2005.08.004.

[88] Nick Freemantle et al. "beta Blockade after myocardial infarction: systematic review and meta regression analysis". In: *BMJ : British Medical Journal* 318.7200 (Apr. 1999), pp. 1730–1737. ISSN: 1468-5833.

[89] Giovanni B Frisoni et al. "The clinical use of structural MRI in Alzheimer disease". In: *Nature reviews. Neurology* 6.2 (Feb. 2010), pp. 67–77. ISSN: 1759-4766. DOI: 10.1038/nrneurol.2009.215.

[90] K.J. Friston et al. *Human Brain Function*. Ed. by Karl J. Friston et al. 2nd ed. Human Brain Function. Elsevier Science, 2004, p. 1144. ISBN: 9780080472959. DOI: ISBN0080472958,9780080472959.

[91] HJ Fu et al. "Amyloid-beta Immunotherapy for Alzheimer's Disease". In: *CNS & neurological disorders drug targets* 9.2 (Apr. 2010), pp. 197–206. ISSN: 1996-3181.

[92] Wen Fu and JackH. Jhamandas. "Role of astrocytic glycolytic metabolism in Alzheimer's disease pathogenesis". English. In: *Biogerontology* 15.6 (2014), pp. 579–586. ISSN: 1389-5729. DOI: 10.1007/s10522-014-9525-0.

[93] R Fujita et al. "Prothymosin-alpha plays a defensive role in retinal ischemia through necrosis and apoptosis inhibition". In: *Cell Death Differ* 16.2 (Nov. 2008), pp. 349–358. ISSN: 1350-9047. DOI: 10.1038/cdd.2008.159.

[94] R. Ganesan et al. "Proteolytic activation of pro-macrophage-stimulating protein by hepsin". In: *Molecular cancer research : MCR* 9.9 (2011), pp. 1175–86. ISSN: 1557-3125 (Electronic) 1541-7786 (Linking). DOI: 10.1158/1541-7786.MCR-11-0004.

[95] Amit X. Garg, Dan Hackam and Marcello Tonelli. "Systematic Review and Meta-analysis: When One Study Is Just not Enough". In: *Clinical Journal of the American Society of Nephrology* 3.1 (2008), pp. 253–260. DOI: 10.2215/CJN.01430307. eprint: http://cjasn.asnjournals.org/content/3/1/253.full.pdf+html.

[96] Maciej Gasior, Michael A Rogawski and Adam L Hartman. "Neuroprotective and disease-modifying effects of the ketogenic diet". In: *Behavioural pharmacology* 17.5-6 (Sept. 2006), pp. 431–439. ISSN: 0955-8810.

[97] R Gentleman et al. *Genefilter: Methods for filtering genes from microarray experiments*. 0. 2011.

[98] Olivier Gevaert et al. "Predicting the prognosis of breast cancer by integrating clinical and microarray data with Bayesian networks". In: *Bioinformatics* 22.14 (2006), e184–e190. DOI: 10.1093/bioinformatics/btl230. eprint: http://bioinformatics.oxfordjournals.org/content/22/14/e184.full.pdf+html.

[99] Zoubin Ghahramani. "Unsupervised Learning". English. In: *Advanced Lectures on Machine Learning*. Ed. by Olivier Bousquet, Ulrike von Luxburg and Gunnar RГhtsch. Vol. 3176. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2004, pp. 72–112. ISBN: 978-3-540-23122-6. DOI: 10.1007/978-3-540-28650-9_5.

[100] Debashis Ghosh et al. "Statistical issues and methods for meta-analysis of microarray data: a case study in prostate cancer". English. In: *Functional & Integrative Genomics* 3.4 (2003), pp. 180–188. ISSN: 1438-793X. DOI: 10.1007/s10142-003-0087-5.

[101] Alfred L Goldberg, Ross Stein and Julian Adams. "New insights into proteasome function: from archaebacteria to drug development". In: *Chemistry & Biology* 2.8 (Aug. 1995), pp. 503–508. ISSN: 1074-5521. DOI: 10.1016/1074-5521(95)90182-5.

[102] Teresa Gomez-Isla et al. "Erratum for Gomez-Isla et al., Profound Loss of Layer II Entorhinal Cortex Neurons Occurs in Very Mild AlzheimerвГЄs Disease". In: *The Journal of Neuroscience* 17.14 (1997), np. eprint: http://www.jneurosci.org/content/17/14/np.full.pdf+html.

[103] Martin Gomez Ravetti, Regina Berretta and Pablo Moscato. "Novel Biomarkers for Prostate Cancer Revealed by (α, β) - k Feature Sets". In: *Foundations of Computational Intelligence Volume 5*. Ed. by Ajith Abraham, Aboul-Ella Hassanien and Vaclav Snasel. Vol. 205. Studies in Computational Intelligence. Springer Berlin Heidelberg, 2009, pp. 149–175. ISBN: 978-3-642-01535-9.

[104] Martin Gomez Ravetti and Pablo Moscato. "Identification of a 5-Protein Biomarker Molecular Signature for Predicting Alzheimer's Disease". In: *PLoS ONE* 3.9 (Sept. 2008), e3111. DOI: 10.1371/journal.pone.0003111.

[105] Martin Gomez Ravetti et al. "Uncovering Molecular Biomarkers That Correlate Cognitive Decline with the Changes of Hippocampus' Gene Expression Profiles in Alzheimer's Disease". In: *PLoS ONE* 5.4 (Apr. 2010), e10153. DOI: 10.1371/journal.pone.0010153.

[106] I.J Goods. "On the Weighted Combination of Significance Tests". In: *Journal of the Royal Statistical Society* 17 (1955), pp. 264–265.

[107] Tobias Gorges et al. "Circulating tumour cells escape from EpCAM-based detection due to epithelial-to-mesenchymal transition". In: *BMC Cancer* 12.1 (2012), p. 178. ISSN: 1471-2407. DOI: 10.1186/1471-2407-12-178.

[108] Ivan Gorlov et al. "Candidate pathways and genes for prostate cancer: a meta-analysis of gene expression data". In: *BMC Medical Genomics* 2.1 (2009), p. 48. ISSN: 1755-8794.

[109] Rajgopal Govindarajan et al. "Impaired Trafficking of Connexins in Androgen-independent Human Prostate Cancer Cell Lines and Its Mitigation by α-Catenin". In: *Journal of Biological Chemistry* 277.51 (2002), pp. 50087–50097. DOI: 10.1074/jbc.M202652200.

[110] Cheryl L. Grady et al. "Evidence from Functional Neuroimaging of a Compensatory Prefrontal Network in Alzheimer's Disease". In: *The Journal of Neuroscience* 23.3 (2003), pp. 986–993. eprint: http://www.jneurosci.org/content/23/3/986.full.pdf+html.

[111] Atle van Beelen Granlund et al. "Whole Genome Gene Expression Meta-Analysis of Inflammatory Bowel Disease Colon Mucosa Demonstrates Lack of Major Differences between Crohn's Disease and Ulcerative Colitis". In: *PLoS ONE* 8.2 (2013), e56818.

[112] Dario Greco et al. "Physiology, Pathology and Relatedness of Human Tissues from Gene Expression Meta-Analysis". In: *PLoS ONE* 3.4 (Feb. 2008), e1880–. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0001880.

[113] J.Timothy Greenamyre and Anne B. Young. "Excitatory amino acids and Alzheimer's disease". In: *Neurobiology of Aging* 10.5 (1989), pp. 593 –602. ISSN: 0197-4580. DOI: http://dx.doi.org/10.1016/0197-4580(89)90143-7.

[114] Prabhjit K. Grewal et al. "Characterization of the LARGE family of putative glycosyltransferases associated with dystroglycanopathies". In: *Glycobiology* 15.10 (2005), pp. 912–923. DOI: 10.1093/glycob/cwi094. eprint: http://glycob.oxfordjournals.org/content/15/10/912.full.pdf+html.

[115] Sam Griffiths-Jones et al. "miRBase: microRNA sequences, targets and gene nomenclature". In: *Nucleic Acids Research* 34.suppl 1 (2006), pp. D140–D144. DOI: 10.1093/nar/gkj112. eprint: http://nar.oxfordjournals.org/content/34/suppl_1/D140.full.pdf+html.

[116] MariaConcetta Gueli and Gennaro Taibi. "Alzheimer's disease: amino acid levels and brain metabolic status". English. In: *Neurological Sciences* 34.9 (2013), pp. 1575–1579. ISSN: 1590-1874. DOI: 10.1007/s10072-013-1289-9.

[117] R. Guerra and D.R. Goldstein. *Meta-analysis and Combining Information in Genetics and Genomics*. Chapman & Hall/CRC Mathematical and Computational Biology. Taylor & Francis, 2009. ISBN: 9781584885221.

[118] Rita J. Guerreiro, Deborah R. Gustafson and John Hardy. "The genetic architecture of Alzheimer's disease: beyond APP, PSENs and APOE". In: *Neurobiology of Aging* 33.3 (2012), pp. 437 –456. ISSN: 0197-4580. DOI: 10.1016/j.neurobiolaging.2010.03.025.

[119] Changyong Guo et al. "Epcam, CD44, and CD49f Distinguish Sphere–Forming Human Prostate Basal Cells from a Subpopulation with Predominant Tubule Initiation Capability". In: *PLoS ONE* 7.4 (Apr. 2012), e34219. DOI: 10.1371/journal.pone.0034219.

[120] J. Guo et al. "HLA-A2-restricted cytotoxic T lymphocyte epitopes from human hepsin as novel targets for prostate cancer immunotherapy". In: *Scandinavian journal of immunology* 78.3 (2013), pp. 248–57. ISSN: 1365-3083 (Electronic) 0300-9475 (Linking). DOI: 10.1111/sji.12083.

[121] Isabelle Guyon and André Elisseeff. "An Introduction to Variable and Feature Selection". In: *J. Mach. Learn. Res.* 3 (Mar. 2003), pp. 1157–1182. ISSN: 1532-4435.

[122] A B Haidich. "Meta-analysis in medical research". In: *Hippokratia* 14.Suppl 1 (Dec. 2010), pp. 29–37. ISSN: 1790-8019.

[123] Chadwick M Hales et al. "Abnormal Gephyrin Immunoreactivity Associated with Alzheimer's Disease Pathologic Changes". In: *Journal of neuropathology and experimental neurology* 72.11 (Nov. 2013), pp. 1009–1015. ISSN: 1554-6578.

[124] Mark A Hall. "Correlation-based feature selection for machine learning". PhD thesis. The University of Waikato, 1999.

[125] Jemila S. Hamid et al. "Data Integration in Genetics and Genomics: Methods and Challenges". In: *Human Genomics and Proteomics* 1.1 (2009), p. 869093. DOI: 10.4061/2009/869093. eprint: http://hgp.sagepub.com/content/1/1/869093.full.pdf+html.

[126] Gerhard Hamilton, Ulrike Olszewski-Hamilton and Gerhard Theyer. "Type I Collagen Synthesis Marker Procollagen I N-Terminal Peptide (PINP) in Prostate Cancer Patients Undergoing Intermittent Androgen Suppression". In: *Cancers* 3.3 (2011), pp. 3601–3609. ISSN: 2072-6694.

[127] Douglas Hanahan and Robert A Weinberg. "The Hallmarks of Cancer". In: *Cell* 100.1 (2000), pp. 57 –70. ISSN: 0092-8674. DOI: 10.1016/S0092-8674(00)81683-9.

[128] K. M. Haram et al. "Gene expression profile of mouse prostate tumors reveals dysregulations in major biological processes and identifies potential murine targets for preclinical development of human prostate cancer therapy". In: *The Prostate* 68.14 (2008), pp. 1517–30. ISSN: 1097-0045 (Electronic) 0270-4137 (Linking). DOI: 10.1002/pros.20803.

[129] J. A. Harasty et al. "Specific temporoparietal gyral atrophy reflects the pattern of language dissolution in Alzheimer's disease". In: *Brain* 122.4 (1999), pp. 675–686. ISSN: 0006-8950. DOI: 10.1093/brain/122.4.675.

[130] Xiaofei He, Deng Cai and Partha Niyogi. "Laplacian score for feature selection". In: *In NIPS*. MIT Press, 2005.

[131] Elizabeth Head and Frederick A Schmitt. "Con: are we ready to translate Alzheimer's disease-modifying therapies to people with down syndrome?" In: *Alzheimer's Research & Therapy* 6.5 (Sept. 2014), pp. 61–61. ISSN: 1758-9193. DOI: 10.1186/s13195-014-0061-6.

[132] L.V. Hedges and I. Olkin. *Statistical Methods for Meta-analysis*. Academic Press, 1985. ISBN: 9780123363817.

[133] Larry V; Hedges. *Fixed effects models*. Ed. by Harris; Cooper, Larry V; Hedges and Jeffrey C; Valentine. New York: Russell Sage Foundation, 1994, pp. 285–300.

[134] Larry V. Hedges and Jack L. Vevea. "Fixed- and random-effects models in meta-analysis." In: *Psychological Methods* 3.4 (1998), pp. 486–504. ISSN: 1939-1463(Electronic);1082-989X(Print). DOI: 10.1037/1082-989X.3.4.486.

[135] Sarah R Heilbronner and Michael L Platt. "Causal evidence of performance monitoring by neurons in posterior cingulate cortex during learning". In: *Neuron* 80.6 (Dec. 2013), pp. 1384–1391. ISSN: 1097-4199. DOI: 10.1016/j.neuron.2013.09.028.

[136] Matthew L. Hemming et al. “Identification of ΓΗΒў-Secretase (BACE1) Substrates Using Quantitative Proteomics”. In: *PLoS ONE* 4.12 (Dec. 2009), e8477–. DOI: 10.1371/journal.pone.0008477.

[137] Samuel T Henderson. “High carbohydrate diets and Alzheimer’s disease”. In: *Medical Hypotheses* 62.5 (2004), pp. 689 –700. ISSN: 0306-9877. DOI: http://dx.doi.org/10.1016/j.mehy.2003.11.028.

[138] Masaaki Hokama et al. “Altered Expression of Diabetes-Related Genes in Alzheimer’s Disease Brains: The Hisayama Study”. In: *Cerebral Cortex* 24.9 (2014), pp. 2476–2488. DOI: 10.1093/cercor/bht101. eprint: http://cercor.oxfordjournals.org/content/24/9/2476.full.pdf+html.

[139] Fangxin Hong and Rainer Breitling. “A comparison of meta-analysis methods for detecting differentially expressed genes in microarray experiments”. In: *Bioinformatics* 24.3 (2008), pp. 374–382.

[140] Fangxin Hong et al. “RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis”. In: *Bioinformatics* 22.22 (2006), pp. 2825–2827. DOI: 10.1093/bioinformatics/btl476. eprint: http://bioinformatics.oxfordjournals.org/content/22/22/2825.full.pdf+html.

[141] Roger Horton et al. “Gene map of the extended human MHC”. In: *Nat Rev Genet* 5.12 (Dec. 2004), pp. 889–899. ISSN: 1471-0056. DOI: 10.1038/nrg1489.

[142] Douglas Hosack et al. “Identifying biological themes within lists of genes with EASE”. In: *Genome Biology* 4.10 (2003). A previous version of this manuscript was made available before peer review at http://genomebiology.com/2003/4/6/P4, R70. ISSN: 1465-6906. DOI: 10.1186/gb-2003-4-10-r70.

[143] Amir M Hossini et al. “Induced pluripotent stem cell-derived neuronal cells from a sporadic AlzheimerвЂ™s disease donor as a model for investigating AD-associated gene regulatory networks”. In: *BMC Genomics* 16.1 (Jan. 2015), p. 84. ISSN: 1471-2164. DOI: 10.1186/s12864-015-1262-5.

[144] Chih-lin Hsieh et al. “A Genome Screen of Families with Multiple Cases of Prostate Cancer: Evidence of Genetic Heterogeneity”. In: *The American Journal of Human Genetics* 69.1 (2001), pp. 148–158. ISSN: 0002-9297. DOI: http://dx.doi.org/10.1086/321281.

[145] Jing Hu et al. “Inhibition of Monocyte Adhesion to Brain-Derived Endothelial Cells by Dual Functional RNA Chimeras”. In: *Mol Ther Nucleic Acids* 3 (Nov. 2014), e209–. DOI: 10.1038/mtna.2014.60.

[146] Pingzhao Hu, CeliaM.T. Greenwood and Joseph Beyene. "Statistical Methods for Meta-Analysis of Microarray Data: A Comparative Study". English. In: *Information Systems Frontiers* 8.1 (2006), pp. 9–20. ISSN: 1387-3326. DOI: 10.1007/s10796-005-6099-z.

[147] Yunping Hu et al. "Syndecan-1-Dependent Suppression of PDK1/Akt/Bad Signaling by Docosahexaenoic Acid Induces Apoptosis in Prostate Cancer". In: *Neoplasia* 12.10 (2010), pp. 826 –836. ISSN: 1476-5586. DOI: 10.1593/neo.10586.

[148] Da Wei Huang, Brad T Sherman and Richard A Lempicki. "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources". In: *Nat. Protocols* 4.1 (Dec. 2008), pp. 44–57. ISSN: 1754-2189. DOI: 10.1038/nprot.2008.211.

[149] Christian Humpel. "Identifying and validating biomarkers for Alzheimer's disease". In: *Trends in Biotechnology* 29.1 (Jan. 2011), pp. 26–32. ISSN: 1879-3096.

[150] MF Huque. "Experiences with meta-analysis in NDA submissions". In: *Proceedings of the Biopharmaceutical Section of the American Statistical Association*. Vol. 2. 1988, pp. 28–33.

[151] S. Inagaki, T. Furuyama and Y. Iwahashi. "Identification of a member of mouse semaphorin family". In: *FEBS Letters* 370.3 (Aug. 1995), pp. 269–272. ISSN: 0014-5793. DOI: 10.1016/0014-5793(95)00850-9.

[152] Shinobu Inagaki et al. "Sema4C, a Transmembrane Semaphorin, Interacts with a Post-synaptic Density Protein, PSD-95". In: *Journal of Biological Chemistry* 276.12 (2001), pp. 9174–9181. DOI: 10.1074/jbc.M009051200. eprint: http://www.jbc.org/content/276/12/9174.full.pdf+html.

[153] Rafael A Irizarry et al. "Multiple-laboratory comparison of microarray platforms". In: *Nat Meth* 2.5 (2005), pp. 345–350. ISSN: 1548-7091.

[154] Masahiko Iwakiri et al. "Changes in hippocampal GABABR1 subunit expression in Alzheimer's patients: association with Braak staging". English. In: *Acta Neuropathologica* 109.5 (2005), pp. 467–474. ISSN: 0001-6322. DOI: 10.1007/s00401-005-0985-9.

[155] Peyman Jafari and Francisco Azuaje. "An assessment of recently published gene expression data analyses: reporting experimental design and statistical factors". In: *BMC Medical Informatics and Decision Making* 6.1 (2006), p. 27. ISSN: 1472-6947. DOI: 10.1186/1472-6947-6-27.

[156] Hongying Jiang et al. "Joint analysis of two microarray gene-expression data sets to select lung adenocarcinoma marker genes." In: *BMC Bioinformatics* 5.1 (2004), p. 81. ISSN: 1471-2105. DOI: 10.1186/1471-2105-5-81.

[157] Ning Jiang et al. "A-Methylacyl-CoA Racemase (AMACR) and Prostate-Cancer Risk: A Meta-Analysis of 4,385 Participants". In: *PLoS ONE* 8.10 (2013), e74386. DOI: 10.1371/journal.pone.0074386.

[158] Thanyaluk Jirapech-Umpai and Stuart Aitken. "Feature selection and classification for microarray data analysis: Evolutionary methods for identifying predictive genes". In: *BMC Bioinformatics* 6.1 (2005), p. 148. ISSN: 1471-2105. DOI: 10.1186/1471-2105-6-148.

[159] Keith A. Johnson et al. "Brain Imaging in Alzheimer Disease". In: *Cold Spring Harbor Perspectives in Medicine* 2.4 (2012). a006213[PII] 22474610[pmid] Cold Spring Harb Perspect Med, a006213. ISSN: 2157-1422. DOI: 10.1101/cshperspect.a006213.

[160] Rory Johnson. "Long non-coding {RNAs} in Huntington's disease neurodegeneration". In: *Neurobiology of Disease* 46.2 (2012). MicroRNAs in Neuropsychiatric Disease, pp. 245 –254. ISSN: 0969-9961. DOI: http://dx.doi.org/10.1016/j.nbd.2011.12.006.

[161] Bethany F. Jones et al. "Differential Regional Atrophy of the Cingulate Gyrus in Alzheimer Disease: A Volumetric MRI Study". In: *Cerebral Cortex* 16.12 (2006), pp. 1701–1708. DOI: 10.1093/cercor/bhj105. eprint: http://cercor.oxfordjournals.org/content/16/12/1701.full.pdf+html.

[162] E.G. Jones. *Cancer Its Cases, Symptoms and Treatment*. B. Jain Publishers (P) Limited, 2003. ISBN: 9788170211426.

[163] T. Juliusdottir et al. "Two-Phase EA/k-NN for Feature Selection and Classification in Cancer Microarray Datasets". In: *Computational Intelligence in Bioinformatics and Computational Biology, 2005. CIBCB '05. Proceedings of the 2005 IEEE Symposium on*. Nov. 2005, pp. 1–8. DOI: 10.1109/CIBCB.2005.1594891.

[164] Eunsung Junn et al. "Repression of Oç-synuclein expression and toxicity by microRNA-7". In: *Proceedings of the National Academy of Sciences of the United States of America* 106.31 (Mar. 2009), pp. 13052–13057. ISSN: 1091-6490. DOI: 10.1073/pnas.0906277106.

[165] I Kadish et al. "Hippocampal and Cognitive Aging across the Lifespan: A Bioenergetic Shift Precedes and Increased Cholesterol Trafficking Parallels Memory Impairment". In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 29.6 (Feb. 2009), pp. 1805–1816. ISSN: 1529-2401. DOI: 10.1523/JNEUROSCI.4599-08.2009.

[166] Minoru Kanehisa and Susumu Goto. "KEGG: Kyoto Encyclopedia of Genes and Genomes". In: *Nucleic Acids Research* 28.1 (2000), pp. 27–30. DOI: 10.1093/nar/28.1.27. eprint: http://nar.oxfordjournals.org/content/28/1/27.full.pdf+html.

[167] Jihong Kang and Serge Rivest. "Lipid Metabolism and Neuroinflammation in Alzheimer's Disease: A Role for Liver X Receptors". In: *Endocrine Reviews* 33.5 (2012). PMID: 22766509, pp. 715–746. DOI: 10.1210/er.2011-1049. eprint: http://dx.doi.org/10.1210/er.2011-1049.

[168] Jing-Qiong Kang, Wangzhen Shen and Robert L. Macdonald. "The GABRG2 Mutation, Q351X, Associated with Generalized Epilepsy with Febrile Seizures Plus, Has Both Loss of Function and Dominant-Negative Suppression". In: *The Journal of Neuroscience* 29.9 (2009), pp. 2845–2856. DOI: 10.1523/JNEUROSCI.4772-08.2009. eprint: http://www.jneurosci.org/content/29/9/2845.full.pdf+html.

[169] Katja Kanninen et al. "Glycosylation changes in Alzheimer's disease as revealed by a proteomic approach". In: *Neuroscience Letters* 367.2 (2004), pp. 235 –240. ISSN: 0304-3940. DOI: http://dx.doi.org/10.1016/j.neulet.2004.06.013.

[170] Pearson Karl. "Report on certain enteric fever inoculation statistics." In: *British Medical Journal* 3 (1904), pp. 1243–1246.

[171] N.K. Kasabov. *Springer Handbook of Bio-Neuro-Informatics*. SpringerLink : Bucher. Springer Berlin Heidelberg, 2013, p. 955. ISBN: 9783642305740.

[172] Menachem Katz, Ido Amit and Yosef Yarden. "Regulation of MAPKs by growth factors and receptor tyrosine kinases". In: *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1773.8 (Aug. 2007), pp. 1161–1176. ISSN: 0167-4889. DOI: 10.1016/j.bbamcr.2007.01.002.

[173] Audrey Kauffmann, Robert Gentleman and Wolfgang Huber. "arrayQualityMetrics–a bioconductor package for quality assessment of microarray data". In: *Bioinformatics* 25.3 (Dec. 2008), pp. 415–416. ISSN: 1460-2059. DOI: 10.1093/bioinformatics/btn647.

[174] James M Kelley, Laura B Hughes and S Louis Bridges. "Does gamma-aminobutyric acid (GABA) influence the development of chronic inflammation in rheumatoid arthritis?" In: *Journal of Neuroinflammation* 5 (Jan. 2008), pp. 1–1. ISSN: 1742-2094. DOI: 10.1186/1742-2094-5-1.

[175] Usman A Khan et al. "Molecular drivers and cortical spread of lateral entorhinal cortex dysfunction in preclinical Alzheimer's disease". In: *Nat Neurosci* 17.2 (Feb. 2014), pp. 304–311. ISSN: 1097-6256.

[176] Eun Kyung Kim and Eui-Ju Choi. "Pathological roles of MAPK signaling pathways in human diseases". In: *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1802.4 (2010), pp. 396 –405. ISSN: 0925-4439. DOI: http://dx.doi.org/10.1016/j.bbadis.2009.12.009.

[177] H. J. Kim et al. "Variants in the HEPSIN gene are associated with susceptibility to prostate cancer". In: *Prostate cancer and prostatic diseases* 15.4 (2012), pp. 353–8. ISSN: 1476-5608 (Electronic) 1365-7852 (Linking). DOI: 10.1038/pcan.2012.17.

[178] Jeong-Min Kim et al. "Identification of Genes Related to Parkinson's Disease Using Expressed Sequence Tags". In: *DNA Research* 13.6 (2007), pp. 275–286. DOI: 10.1093/dnares/dsl016. eprint: http://dnaresearch.oxfordjournals.org/content/13/6/275.full.pdf+html.

[179] Jungsu Kim, Jacob M Basak and David M Holtzman. "The Role of Apolipoprotein E in Alzheimer's Disease". In: *Neuron* 63.3 (Aug. 2009), pp. 287–303. ISSN: 1097-4199. DOI: 10.1016/j.neuron.2009.06.026.

[180] A.R. Kimmel and B. Oliver. *DNA Microarrays Part A: Array Platforms & Wet-Bench Protocols: Array Platforms & Wet-Bench Protocols*. Methods in enzymology pt. 1. Elsevier Science, 2011. ISBN: 9780080464657.

[181] Eric W Klee et al. "Candidate Serum Biomarkers for Prostate Adenocarcinoma Identified by mRNA Differences in Prostate Tissue and Verified with Protein Measurements in Tissue and Blood." In: *Clinical chemistry* 58.3 (Jan. 2012), pp. 599–609. ISSN: 1530-8561. DOI: 10.1373/clinchem.2011.171637.

[182] Morgan L Kleiber et al. "Neurodevelopmental alcohol exposure elicits long-term changes to gene expression that alter distinct molecular pathways dependent on timing of exposure". In: *Journal of Neurodevelopmental Disorders* 5.1 (Feb. 2013), pp. 6–6. ISSN: 1866-1955.

[183] Ji-Ae Ko et al. "Requirement of the transmembrane semaphorin Sema4C for myogenic differentiation". In: *FEBS Letters* 579.10 (Apr. 2005), pp. 2236–2242. ISSN: 0014-5793. DOI: 10.1016/j.febslet.2005.03.022.

[184] Nikolay Kolesnikov et al. "ArrayExpress update–simplifying data submissions". In: *Nucleic acids research* 43.Database issue (2015), pp. D1113–6. ISSN: 0305-1048. DOI: 10.1093/nar/gku1057.

[185] Wei Kong et al. "Independent component analysis of Alzheimer's DNA microarray gene expression data". In: *Molecular Neurodegeneration* 4 (Jan. 2009), pp. 5–5. ISSN: 1750-1326. DOI: 10.1186/1750-1326-4-5.

[186] Jens Ekkehard Konig et al. "Analysis of the inflammatory network in benign prostate hyperplasia and prostate cancer". In: *The Prostate* 58.2 (2004), pp. 121–129. ISSN: 1097-0045. DOI: 10.1002/pros.10317.

[187] Ravi Kothapalli et al. "Microarray results: how accurate are they?" In: *BMC Bioinformatics* 3.1 (2002), p. 22. ISSN: 1471-2105.

[188] Jennifer R. Kowalski and Peter Juo. "The Role of Deubiquitinating Enzymes in Synaptic Function and Nervous System Diseases". In: *Neural Plasticity* 2012 (2012), p. 13. DOI: 10.1155/2012/892749.

[189] James A. Koziol and Michael D. Perlman. "Combining Independent Chi-Squared Tests". In: *Journal of the American Statistical Association* 73.364 (Dec. 1978), pp. 753–763. ISSN: 01621459. DOI: 10.2307/2286276.

[190] Kristin R Krueger et al. "Social engagement and cognitive function in old age". In: *Experimental aging research* 35.1 (2009), pp. 45–60. ISSN: 1096-4657. DOI: 10.1080/03610730802545028.

[191] Winston Patrick Kuo et al. "Analysis of matched mRNA measurements from two different microarray technologies". In: *Bioinformatics* 18.3 (2002), pp. 405–412. DOI: 10.1093/bioinformatics/18.3.405. eprint: http://bioinformatics.oxfordjournals.org/content/18/3/405.full.pdf+html.

[192] Pak Shing Kwan et al. "Daxx regulates mitotic progression and prostate cancer predisposition". In: *Carcinogenesis* 34.4 (2013), pp. 750–759. DOI: 10.1093/carcin/bgs391.

[193] Y. H. Lai et al. "SOX4 interacts with plakoglobin in a Wnt3a-dependent manner in prostate cancer cells". In: *BMC cell biology* 12 (2011), p. 50. ISSN: 1471-2121 (Electronic) 1471-2121 (Linking). DOI: 10.1186/1471-2121-12-50.

[194] Jean-Charles Lambert et al. "Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease". In: *Nature genetics* 45.12 (Oct. 2013), pp. 1452–1458. ISSN: 1546-1718. DOI: 10.1038/ng.2802.

[195] J. Lapointe et al. "Gene expression profiling identifies clinically relevant subtypes of prostate cancer". In: *Proc Natl Acad Sci U S A* 101 (2004), pp. 811–6–. ISSN: 0027-8424 (Print) 0027-8424 (Linking).

[196] H Le-Niculescu et al. "Convergent functional genomics of anxiety disorders: translational identification of genes, biomarkers, pathways and mechanisms". In: *Transl Psychiatry* 1 (May 2011), e9–. DOI: 10.1038/tp.2011.9.

[197] Candy S Lee et al. "Loss of nuclear factor E2-related factor 1 in the brain leads to dysregulation of proteasome gene expression and neurodegeneration". In: *Proceedings of the National Academy of Sciences of the United States of America* 108.20 (May 2011), pp. 8408–8413. ISSN: 1091-6490. DOI: 10.1073/pnas.1019209108.

[198] Moonhee Lee et al. "Acidic Fibroblast Growth Factor (FGF) Potentiates Glial-mediated Neurotoxicity by Activating FGFR2 IIIb Protein". In: *The Journal of Biological Chemistry* 286.48 (Sept. 2011), pp. 41230–41245. ISSN: 1083-351X. DOI: 10.1074/jbc.M111.270470.

[199] Robert Leech, Rodrigo Braga and David J. Sharp. "Echoes of the Brain within the Posterior Cingulate Cortex". In: *The Journal of Neuroscience* 32.1 (2012), pp. 215–222. DOI: 10.1523/jneurosci.3689-11.2012.

[200] Robert Leech and David J. Sharp. "The role of the posterior cingulate cortex in cognition and disease". In: *Brain* 137 (2013), pp. 12–32. DOI: http://dx.doi.org/10.1093/brain/awt162.

[201] Cynthia A Lemere and Eliezer Masliah. "Can Alzheimer disease be prevented by amyloid-beta immunotherapy?" In: *Nature reviews. Neurology* 6.2 (Feb. 2010), pp. 108–119. ISSN: 1759-4766. DOI: 10.1038/nrneurol.2009.219.

[202] Nathan E Lewis et al. "Large-scale in silico modeling of metabolic interactions between cell types in the human brain". In: *Nat Biotech* 28.12 (Dec. 2010), pp. 1279–1285. ISSN: 1087-0156. DOI: 10.1038/nbt.1711.

[203] Jia Li and George C. Tseng. "An adaptively weighted statistic for detecting differential gene expression when combining multiple transcriptomic studies". In: *Ann. Appl. Stat.* 5.2A (June 2011), pp. 994–1019. DOI: 10.1214/10-AOAS393.

[204] Leping Li et al. "Gene selection for sample classification based on gene expression data: study of sensitivity to choice of parameters of the GA/KNN method". In: *Bioinformatics* 17.12 (2001), pp. 1131–1142. DOI: 10.1093/bioinformatics/17.12.1131. eprint: http://bioinformatics.oxfordjournals.org/content/17/12/1131.full.pdf+html.

[205] Quefeng Li et al. "Meta-analysis based variable selection for gene expression data". In: *Biometrics* 70.4 (2014), pp. 872–880. ISSN: 1541-0420. DOI: 10.1111/biom.12213.

[206] Rile Li et al. "Neural cell adhesion molecule is upregulated in nerves with prostate cancer invasion". In: *Human Pathology* 34.5 (2003), pp. 457–461. ISSN: 0046-8177. DOI: http://dx.doi.org/10.1016/S0046-8177(03)00084-4.

[207] Xin Li and Richard W. Carthew. "A microRNA Mediates EGF Receptor Signaling and Promotes Photoreceptor Differentiation in the Drosophila Eye". In: *Cell* 123.7 (2005), pp. 1267–1277. ISSN: 0092-8674. DOI: http://dx.doi.org/10.1016/j.cell.2005.10.040.

[208] Dapeng Liang et al. "Concerted Perturbation Observed in a Hub Network in Alzheimer's Disease". In: *PLoS ONE* 7.7 (July 2012), e40498. DOI: 10.1371/journal.pone.0040498.

[209] Winnie S. Liang et al. "Altered neuronal gene expression in brain regions differentially affected by Alzheimer's disease: a reference data set". In: *Physiological genomics* 33.2 (2008). 18270320[pmid] Physiol Genomics, pp. 240–256. ISSN: 1094-8341 1531-2267. DOI: 10.1152/physiolgenomics.00242.2007.

[210] Winnie S. Liang et al. "Alzheimer's disease is associated with reduced expression of energy metabolism genes in posterior cingulate neurons". In: *Proceedings of the National Academy of Sciences of the United States of America* 105.11 (2008). 9616[PII] 18332434[pmid] Proc Natl Acad Sci U S A, pp. 4441–4446. ISSN: 0027-8424 1091-6490. DOI: 10.1073/pnas.0709259105.

[211] Stuart A. Lipton and Paul A. Rosenberg. "Excitatory Amino Acids as a Final Common Pathway for Neurologic Disorders". In: *New England Journal of Medicine* 330.9 (1994). PMID: 7905600, pp. 613–622. DOI: 10.1056/NEJM199403033300907. eprint: http://dx.doi.org/10.1056/NEJM199403033300907.

[212] Fei Liu et al. "Role of glycosylation in hyperphosphorylation of tau in Alzheimer's disease". In: *FEBS Letters* 512 (2002), pp. 101 –106. ISSN: 0014-5793. DOI: http://dx.doi.org/10.1016/S0014-5793(02)02228-7.

[213] Guoying Liu et al. "NetAffx: Affymetrix probesets and annotations". In: *Nucleic Acids Research* 31.1 (Oct. 2002), pp. 82–86. ISSN: 1362-4962.

[214] Li Liu et al. "Trans-Synaptic Spread of Tau Pathology In Vivo". In: *PLoS ONE* 7.2 (Feb. 2012), e31302. DOI: 10.1371/journal.pone.0031302.

[215] Timothy Liu et al. "Transcriptional signaling pathways inversely regulated in Alzheimer's disease and glioblastoma multiform". In: *Scientific Reports* 3 (Dec. 2013), pp. 3467–. DOI: 10.1038/srep03467.

[216] Zhi-Ping Liu et al. "Identifying dysfunctional crosstalk of pathways in various regions of Alzheimer's disease brains". In: *BMC Systems Biology* 4.Suppl 2 (2010), S11. ISSN: 1752-0509. DOI: 10.1186/1752-0509-4-S2-S11.

[217] Matthias W. Lorenz et al. "Prediction of Clinical Cardiovascular Events With Carotid Intima-Media Thickness: A Systematic Review and Meta-Analysis". In: *Circulation* 115.4 (2007), pp. 459–467. DOI: 10.1161/CIRCULATIONAHA.106.628875. eprint: http://circ.ahajournals.org/content/115/4/459.full.pdf+html.

[218] Helena R Lourenco, Olivier C Martin and Thomas Stutzle. *Iterated local search*. Springer, 2003.

[219] C. Luo et al. "Gamma-aminobutyric acid (GABA)-B receptor 1 in cerebellar cortex of essential tremor". In: *Journal of Clinical Neuroscience* 19.6 (2012), pp. 920 –921. ISSN: 0967-5868. DOI: http://dx.doi.org/10.1016/j.jocn.2011.11.001.

[220] Jun Luo et al. "Oç-Methylacyl-CoA Racemase: A New Molecular Marker for Prostate Cancer". In: *Cancer Research* 62.8 (2002), pp. 2220–2226.

[221] Lara Lusa, R. Gentleman and M. Ruschhaupt. *GeneMeta: MetaAnalysis for High Throughput Experiments*. R package version 1.36.0.

[222] Donald M. Lyall et al. "Alzheimer's Disease Susceptibility Genes APOE and TOMM40, and Hippocampal Volumes in the Lothian Birth Cohort 1936". In: *PLoS ONE* 8.11 (Nov. 2013), e80513. DOI: 10.1371/journal.pone.0080513.

[223] Gary H Lyman and Nicole M Kuderer. "The strengths and limitations of meta-analyses based on aggregate data". In: *BMC Medical Research Methodology* 5 (Apr. 2005), pp. 14–14. ISSN: 1471-2288. DOI: 10.1186/1471-2288-5-14.

[224] Anne Maass et al. "Laminar activity in the hippocampus and entorhinal cortex related to novelty and episodic encoding". In: *Nat Commun* 5 (2014), p. 5547. DOI: 10.1038/ncomms6547.

[225] Olivier C Maes et al. "MicroRNA: Implications for Alzheimer Disease and other Human CNS Disorders". In: *Current Genomics* 10.3 (Mar. 2009), pp. 154–168. ISSN: 1875-5488. DOI: 10.2174/138920209788185252.

[226] Hiroshi Mamitsuka. "Selecting features in microarray classification using {ROC} curves". In: *Pattern Recognition* 39.12 (2006). Bioinformatics, pp. 2393 –2404. ISSN: 0031-3203. DOI: 10.1016/j.patcog.2006.07.010.

[227] Emily N. Manning et al. "APOE Ox4 Is Associated with Disproportionate Progressive Hippocampal Atrophy in AD". In: *PLoS ONE* 9.5 (May 2014), e97608. DOI: 10.1371/journal.pone.0097608.

[228] Nathan Mantel and William Haenszel. "Statistical Aspects of the Analysis of Data From Retrospective Studies of Disease". In: *Journal of the National Cancer Institute* 22.4 (1959), pp. 719–748. DOI: 10.1093/jnci/22.4.719. eprint: http://jnci.oxfordjournals.org/content/22/4/719.full.pdf+html.

[229] M. Martinez et al. "Amino acid concentrations in cerebrospinal fluid and serum in Alzheimer's disease and vascular dementia". English. In: *Journal of Neural Transmission - Parkinson's Disease and Dementia Section* 6.1 (1993), pp. 1–9. ISSN: 0936-3076. DOI: 10.1007/BF02252617.

[230] Tristan Mary-Huard, Franck Picard and Stephane Robin. "Introduction to statistical methods for microarray data analysis". In: *Mathematical and Computational Methods in Biology* (2006), pp. 56–126.

[231] P. Massoner et al. "EpCAM is overexpressed in local and metastatic prostate cancer, suppressed by chemotherapy and modulated by MET-associated miRNA-200c/205". In: *British journal of cancer* 111.5 (2014), pp. 955–64. ISSN: 1532-1827 (Electronic) 0007-0920 (Linking). DOI: 10.1038/bjc.2014.366.

[232] Mark P. Mattson. "Pathways towards and away from Alzheimer's disease". In: *Nature* 430.7000 (Aug. 2004), pp. 631–639. ISSN: 0028-0836. DOI: 10.1038/nature02621.

[233] Alexandre Mendes, RodneyJ. Scott and Pablo Moscato. "Microarrays–Identifying Molecular Portraits for Prostate Tumors with Different Gleason Patterns". English. In: *Clinical Bioinformatics*. Ed. by RonaldJ.A. Trent. Vol. 141. Methods in Molecular Medicine. Humana Press, 2008, pp. 131–151. ISBN: 978-1-58829-791-4. DOI: 10.1007/978-1-60327-148-6_8.

[234] Alexandre de Mendonca. "Rethinking Alzheimer's Disease". In: *Frontiers in Neurology* 3 (Mar. 2012), p. 45. ISSN: 1664-2295. DOI: 10.3389/fneur.2012.00045.

[235] Tim R. Mercer, Marcel E. Dinger and John S. Mattick. "Long non-coding RNAs: insights into functions". In: *Nat Rev Genet* 10.3 (Mar. 2009), pp. 155–159. ISSN: 1471-0056.

[236] Tim R Mercer et al. "Specific expression of long noncoding RNAs in the mouse brain". In: *Proceedings of the National Academy of Sciences of the United States of America* 105.2 (July 2007), pp. 716–721. ISSN: 1091-6490.

[237] Tim Mercer et al. "Long noncoding RNAs in neuronal-glial fate specification and oligodendrocyte lineage maturation". In: *BMC Neuroscience* 11.1 (2010), p. 14. ISSN: 1471-2202. DOI: 10.1186/1471-2202-11-14.

[238] Emilio Merlo Pich et al. "Imaging as a biomarker in drug discovery for Alzheimer's disease: is MRI a suitable technology?" In: *Alzheimer's Research & Therapy* 6.4 (2014), p. 51. ISSN: 1758-9193. DOI: 10.1186/alzrt276.

[239] Daniel E. Michele et al. "Post-translational disruption of dystroglycan-ligand interactions in congenital muscular dystrophies". In: *Nature* 418.6896 (July 2002), pp. 417–421. ISSN: 0028-0836. DOI: 10.1038/nature00837.

[240] Melissa B Miller and Yi-Wei Tang. "Basic Concepts of Microarrays and Potential Applications in Clinical Microbiology". In: *Clinical Microbiology Reviews* 22.4 (Oct. 2009), pp. 611–633. ISSN: 1098-6618.

[241] E. A. Milward et al. "Beyond Statistics: A New Combinatorial Approach to Identifying Biomarker Panels for the Early Detection and Diagnosis of Alzheimer's Disease". In: *Journal of Alzheimers Disease* 39.1 (2014), pp. 211–217. ISSN: 1387-2877. DOI: Doi10.3233/Jad-131424.

[242] Junxia Min et al. "An oncogene-tumor suppressor cascade drives metastatic prostate cancer by coordinately activating Ras and nuclear factor-[kappa]B". In: *Nat Med* 16.3 (Mar. 2010), pp. 286–294. ISSN: 1078-8956. DOI: 10.1038/nm.2100.

[243] V H Minces et al. "The role of visual cortex acetylcholine in learning to discriminate temporally modulated visual stimuli". In: *Frontiers in Behavioral Neuroscience* 7 (Feb. 2013), pp. 16–. ISSN: 1662-5153. DOI: 10.3389/fnbeh.2013.00016.

[244] David Moher et al. "Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement". In: *PLoS Med* 6.7 (July 2009), e1000097. DOI: 10.1371/journal.pmed.1000097.

[245] Caroline Moreau-Fauvarque et al. "The Transmembrane Semaphorin Sema4D/CD100, an Inhibitor of Axonal Growth, Is Expressed on Oligodendrocytes and Up-regulated after CNS Lesion". In: *The Journal of Neuroscience* 23.27 (2003), pp. 9229–9239. eprint: http://www.jneurosci.org/content/23/27/9229.full.pdf+html.

[246] C. S. Moreno. "The Sex-determining region Y-box 4 and homeobox C6 transcriptional networks in prostate cancer progression: crosstalk with the Wnt, Notch, and PI3K pathways". In: *The American journal of pathology* 176.2 (2010), pp. 518–27. ISSN: 1525-2191 (Electronic) 0002-9440 (Linking). DOI: 10.2353/ajpath.2010.090657.

[247] Martha Clare Morris et al. "DIetary fats and the risk of incident alzheimer disease". In: *Archives of Neurology* 60.2 (2003), pp. 194–200. DOI: 10.1001/archneur.60.2.194. eprint: /data/Journals/NEUR/13502/NOC20214.pdf.

[248] P. Moscato, A. Mendes and R. Berretta. "Benchmarking a memetic algorithm for ordering microarray data". In: *Biosystems* 88.1-2 (2007), pp. 56–75. ISSN: 0303-2647 (Print) 0303-2647 (Linking). DOI: 10.1016/j.biosystems.2006.04.005.

[249] Pablo Moscato. "On evolution, search, optimization, genetic algorithms and martial arts: Towards memetic algorithms". In: *Caltech Concurrent Computation Program* C3P Report (1989), 826:1989.

[250] Pablo Moscato and Carlos Cotta. "A gentle introduction to memetic algorithms". In: *Handbook of metaheuristics*. Springer, 2003, pp. 105–144.

[251] Pablo Moscato, Carlos Cotta and Alexandre Mendes. "Memetic algorithms". In: *New optimization techniques in engineering*. Springer, 2004, pp. 53–85.

[252] Pablo Moscato et al. "The Electronic Primaries : Predicting the U.S. Presidency Using Feature Selection with Safe Data Reduction". In: *Proceedings of the Twenty-eighth Australasian Conference on Computer Science – Volume 38*. ACSC '05. Newcastle, Australia: Australian Computer Society, Inc., 2005, pp. 371–379. ISBN: 1-920-68220-1.

[253] Lisa Mosconi et al. "Pre-Clinical Detection of Alzheimer?s Disease Using FDG-PET, with or without Amyloid Imaging". In: *Journal of Alzheimer's disease : JAD* 20.3 (2010). 20182025[pmid] J Alzheimers Dis, pp. 843–854. ISSN: 1387-2877 1875-8908. DOI: 10.3233/JAD-2010-091504.

[254] Frederick Mosteller and R. A. Fisher. "Questions and Answers". In: *The American Statistician* 2.5 (Oct. 1948), pp. 30–31. ISSN: 00031305. DOI: 10.2307/2681650.

[255] Yangling Mu and Fred Gage. "Adult hippocampal neurogenesis and its role in Alzheimer's disease". In: *Molecular Neurodegeneration* 6.1 (2011), p. 85. ISSN: 1750-1326. DOI: 10.1186/1750-1326-6-85.

[256] Sayan Mukherjee et al. "Estimating dataset size requirements for classifying DNA Microarray data". In: *Journal of Computational Biology* 10 (2003), pp. 119–142.

[257] H.J. Muller and T. Roder. *Microarrays*. Elsevier Academic Press, 2006, p. 117.

[258] R. Mulligan, M. Van der Linden and A.C. Juillerat. *The Clinical Management of Early Alzheimer's Disease: A Handbook*. Taylor & Francis, 2003. ISBN: 9781135665869.

[259] Lenka Munoz and Alaina J. Ammit. "Targeting p38 MAPK pathway for the treatment of Alzheimer's disease". In: *Neuropharmacology* 58.3 (2010), pp. 561 –568. ISSN: 0028-3908. DOI: http://dx.doi.org/10.1016/j.neuropharm.2009.11.010.

[260] Madan L Nagpal, Yue Chen and Tu Lin. "Effects of overexpression of CXCL10 (cytokine-responsive gene-2) on MA-10 mouse Leydig tumor cell steroidogenesis and proliferation". In: *Journal of Endocrinology* 183.3 (2004), pp. 585–594. DOI: 10.1677/joe.1.05795.

[261] Madan L. Nagpal, Jeffrey Davis and Tu Lin. "Overexpression of CXCL10 in human prostate LNCaP cells activates its receptor (CXCR3) expression and inhibits cell proliferation". In: *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1762.9 (2006), pp. 811–818. ISSN: 0925-4439. DOI: http://dx.doi.org/10.1016/j.bbadis.2006.06.017.

[262] Ferrante Neri, Carlos Cotta and Pablo Moscato. *Handbook of memetic algorithms*. Ed. by Heidelberg. Vol. 379. Springer, 2012.

[263] Jordan T. Newington, Richard A. Harris and Robert C. Cumming. "Re-evaluating Metabolism in Alzheimer's Disease from the Perspective of the Astrocyte-Neuron Lactate Shuttle Model". In: *Journal of Neurodegenerative Diseases* 2013 (2013), pp. 1–13. ISSN: 2090-858X. DOI: 10.1155/2013/234572.

[264] Patsopoulos Nikolaos A, Analatos Apostolos A and Ioannidis John P. A. "Relative citation impact of various study designs in the health sciences". In: *JAMA* 293.19 (2005), pp. 2362–2366. DOI: 10.1001/jama.293.19.2362. eprint: /data/Journals/JAMA/4976/JOC50032.pdf.

[265] Sharon-Lise T. Normand. "Meta-analysis: formulating, evaluating, combining, and reporting". In: *Statistics in Medicine* 18.3 (1999), pp. 321–359. ISSN: 1097-0258. DOI: 10.1002/(SICI)1097-0258(19990215)18:3<321::AID-SIM28>3.0.CO;2-P.

[266] Salvatore Oddo. "The ubiquitin-proteasome system in Alzheimer's disease". In: *Journal of Cellular and Molecular Medicine* 12.2 (Feb. 2008), pp. 363–373. ISSN: 1582-4934. DOI: 10.1111/j.1582-4934.2008.00276.x.

[267] Hiroko Ogata-Kawata et al. "Circulating Exosomal microRNAs as Biomarkers of Colon Cancer". In: *PLoS ONE* 9.4 (Apr. 2014), e92921. DOI: 10.1371/journal.pone.0092921.

[268] Yusuke Oji et al. "The translation elongation factor eEF2 is a novel tumor-associated antigen overexpressed in various types of cancers". In: *International Journal of Oncology* 44.5 (Dec. 2014), pp. 1461–1469. ISSN: 1791-2423. DOI: 10.3892/ijo.2014.2318.

[269] C. H. Ooi and Patrick Tan. "Genetic algorithms applied to multi-class prediction for the analysis of gene expression data". In: *Bioinformatics* 19.1 (2003), pp. 37–44. DOI: 10.1093/bioinformatics/19.1.37. eprint: http://bioinformatics.oxfordjournals.org/content/19/1/37.full.pdf+html.

[270] Carlos M Opazo, Mark A Greenough and Ashley I Bush. "Copper: from neurotransmission to neuroproteostasis". In: *Frontiers in Aging Neuroscience* 6 (June 2014), pp. 143–. ISSN: 1663-4365. DOI: 10.3389/fnagi.2014.00143.

[271] Marie Orre et al. "Reactive glia show increased immunoproteasome activity in Alzheimer's disease". In: *Brain* 136.5 (2013), pp. 1415–1431. ISSN: 0006-8950. DOI: 10.1093/brain/awt083.

[272] A. M. Palmer. "The activity of the pentose phosphate pathway is increased in response to oxidative stress in Alzheimer's disease". English. In: *Journal of Neural Transmission* 106.3-4 (1999), pp. 317–328. ISSN: 0300-9564. DOI: 10.1007/s007020050161.

[273] Roberta Paolinelli et al. "Wnt Activation of Immortalized Brain Endothelial Cells as a Tool for Generating a Standardized Model of the Blood Brain Barrier In Vitro". In: *PLoS ONE* 8.8 (June 2013), e70233. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0070233.

[274] Giorgos L. Papadopoulos et al. "The database of experimentally supported targets: a functional update of TarBase". In: *Nucleic Acids Research* 37.suppl 1 (2009), pp. D155–D158. DOI: 10.1093/nar/gkn809. eprint: http://nar.oxfordjournals.org/content/37/suppl_1/D155.full.pdf+html.

[275] Arnaldo Parra-Damas et al. "Crtc1 Activates a Transcriptional Program Deregulated at Early Alzheimer's Disease-Related Stages". In: *The Journal of Neuroscience* 34.17 (2014), pp. 5776–5787. DOI: 10.1523/JNEUROSCI.5288-13.2014. eprint: http://www.jneurosci.org/content/34/17/5776.full.pdf+html.

[276] Nilay V. Patel et al. "Caloric restriction attenuates AOÿ-deposition in Alzheimer transgenic models". In: *Neurobiology of Aging* 26.7 (2005), pp. 995 –1000. ISSN: 0197-4580. DOI: 10.1016/j.neurobiolaging.2004.09.014.

[277] Mateus Rocha de Paula et al. "Differences in Abundances of Cell-Signalling Proteins in Blood Reveal Novel Biomarkers for Early Detection Of Clinical Alzheimer's Disease". In: *PLoS ONE* 6.3 (2011). 10-PONE-RA-19568[PII] 21479255[pmid] PLoS One, e17481. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0017481.

[278] Karl Pearson. "On a Method of Determining Whether a Sample of Size n Supposed to Have Been Drawn from a Parent Population Having a Known Probability Integral has Probably Been Drawn at Random". In: *Biometrika* 25.3/4 (Dec. 1933), pp. 379–410. ISSN: 00063444.

[279] Alessandra Pecorelli et al. "Genes Related to Mitochondrial Functions, Protein Degradation, and Chromatin Folding Are Differentially Expressed in Lymphomonocytes of Rett Syndrome Patients". In: *Mediators of Inflammation* 2013 (2013), p. 18. DOI: 10.1155/2013/137629.

[280] Sarah T Pendlebury and Peter M Rothwell. "Prevalence, incidence, and factors associated with pre-stroke and post-stroke dementia: a systematic review and meta-analysis". In: *The Lancet Neurology* 8.11 (2009), pp. 1006 –1018. ISSN: 1474-4422. DOI: http://dx.doi.org/10.1016/S1474-4422(09)70236-4.

[281] Stephan Persengiev, Ivanela Kondova and Ronald Bontrop. "Insights on the functional interactions between miRNAs and copy number variations in the aging brain". In: *Frontiers in Molecular Neuroscience* 6 (Sept. 2013), p. 32. ISSN: 1662-5099. DOI: 10.3389/fnmol.2013.00032.

[282] Sharon R Pine and Wenyu Liu. "Asymmetric cell division and template DNA co-segregation in cancer stem cells". In: *Frontiers in Oncology* 4.226 (2014). ISSN: 2234-943X. DOI: 10.3389/fonc.2014.00226.

[283] J.G. Pratt et al. *Extra-sensory Perception After Sixty Years: A Critical Appraisal of the Research in Extra-sensory Perception*. H. Holt and Company, 1940.

[284] Nisha Puthiyedth et al. "A New Combinatorial Optimization Approach for Integrated Feature Selection Using Different Datasets: A Prostate Cancer Transcriptomic Study". In: *PLoS ONE* 10.6 (June 2015), e0127702. DOI: 10.1371/journal.pone.0127702.

[285] Nisha Puthiyedth et al. "Identification of Differentially Expressed Genes through Integrated Study of Alzheimer's Disease Affected Brain Regions". In: *PLoS ONE* 11.4 (Apr. 2016), pp. 1–29. DOI: 10.1371/journal.pone.0152342.

[286] M. Qi et al. "ERG rearrangement is associated with prostate cancer-related death in Chinese prostate cancer patients". In: *PloS one* 9.2 (2014), e84959. ISSN: 1932-6203 (Electronic) 1932-6203 (Linking). DOI: 10.1371/journal.pone.0084959.

[287] Irfan A Qureshi and Mark F Mehler. "Emerging roles of non-coding RNAs in brain evolution, development, plasticity and disease". In: *Nature reviews. Neuroscience* 13.8 (July 2012), pp. 528–541. ISSN: 1471-0048.

[288] Victoria A. Rafalski et al. "Expansion of oligodendrocyte progenitor cells following SIRT1 inactivation in the adult brain". In: *Nat Cell Biol* 15.6 (June 2013), pp. 614–624. ISSN: 1465-7392. DOI: 10.1038/ncb2735.

[289] Kammei Rai et al. "Liposomal Delivery of MicroRNA-7b Expressing Plasmid Overcomes Epidermal Growth Factor Receptor Tyrosine Kinase Inhibitor-Resistance in Lung Cancer Cells". In: *Molecular Cancer Therapeutics* 10.9 (2011), pp. 1720–1727. DOI: 10.1158/1535-7163.MCT-11-0220. eprint: http://mct.aacrjournals.org/content/10/9/1720.full.pdf+html.

[290] Adaikalavan Ramasamy et al. "Key Issues in Conducting a Meta-Analysis of Gene Expression Microarray Datasets". In: *PLoS Medicine* 5.9 (Sept. 2008), e184. ISSN: 1549-1676. DOI: 10.1371/journal.pmed.0050184.

[291] Monika Ray and Weixiong Zhang. "Analysis of Alzheimer's disease severity across brain regions by topological analysis of gene co-expression networks". In: *BMC Systems Biology* 4.1 (2010), p. 136. ISSN: 1752-0509. DOI: 10.1186/1752-0509-4-136.

[292] Sandip Ray et al. "Classification and prediction of clinical Alzheimer's diagnosis based on plasma signaling proteins". In: *Nat Med* 13.11 (2007). 10.1038/nm1653, pp. 1359–1362. ISSN: 1078-8956. DOI: http://www.nature.com/nm/journal/v13/n11/suppinfo/nm1653_S1.html.

[293] Tove Reinertsen et al. "Gene expressional changes in prostate fibroblasts from cancerous tissue". In: *APMIS* 120.7 (2012), pp. 558–571. ISSN: 1600-0463. DOI: 10.1111/j.1600-0463.2011.02865.x.

[294] Daniel R. Rhodes et al. "Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression". In: *Proceedings of the National Academy of Sciences of the United States of America* 101.25 (2004), pp. 9309–9314.

[295] Daniel R. Rhodes et al. "Meta-Analysis of Microarrays: Interstudy Validation of Gene Expression Profiles Reveals Pathway Dysregulation in Prostate Cancer". In: *Cancer Research* 62.15 (2002), pp. 4427–4433. eprint: http://cancerres.aacrjournals.org/content/62/15/4427.full.pdf+html.

[296] Beat M Riederer et al. "The role of the ubiquitin proteasome system in Alzheimer's disease". In: *Experimental Biology and Medicine* 236.3 (2011), pp. 268–276. DOI: 10.1258/ebm.2010.010327. eprint: http://ebm.sagepub.com/content/236/3/268.full.pdf+html.

[297] R. Riesenberg et al. "Lysis of prostate carcinoma cells by trifunctional bispecific antibodies (alpha EpCAM x alpha CD3)". In: *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* 49.7 (2001), pp. 911–7. ISSN: 0022-1554 (Print) 0022-1554 (Linking).

[298] Gorica Ristic, Wei-Ling Tsou and Sokol V Todi. "An optimal ubiquitin-proteasome pathway in the nervous system: the role of deubiquitinating enzymes". In: *Frontiers in Molecular Neuroscience* 7 (July 2014), pp. 72–. ISSN: 1662-5099. DOI: 10.3389/fnmol.2014.00072.

[299] Matthew Ritchie et al. "Empirical array quality weights in the analysis of microarray data". In: *BMC Bioinformatics* 7.1 (2006), p. 261. ISSN: 1471-2105. DOI: 10.1186/1471-2105-7-261.

[300] C. Riveros et al. "A transcription factor map as revealed by a genome-wide gene expression analysis of whole-blood mRNA transcriptome in multiple sclerosis". In: *PLoS One* 5.12 (2010), e14176. ISSN: 1932-6203 (Electronic) 1932-6203 (Linking). DOI: 10.1371/journal.pone.0014176.

[301] Richard J Roberts. "PubMed Central: The GenBank of the published literature". In: *Proceedings of the National Academy of Sciences of the United States of America* 98.2 (Jan. 2001), pp. 381–382. ISSN: 1091-6490.

[302] M. A. Rubin et al. "A-methylacyl coenzyme a racemase as a tissue biomarker for prostate cancer". In: *JAMA* 287.13 (2002). 10.1001/jama.287.13.1662, pp. 1662–1670. ISSN: 0098-7484. DOI: 10.1001/jama.287.13.1662.

[303] Maxim Rybalov et al. "PSMA, EpCAM, VEGF and GRPR as Imaging Targets in Locally Recurrent Prostate Cancer after Radiotherapy". In: *International Journal of Molecular Sciences* 15.4 (Mar. 2014), pp. 6046–6061. ISSN: 1422-0067.

[304] Henry. Sacks, Thomas C.. Chalmers and Harry. Smith Jr. "Randomized versus historical controls for clinical trials". In: *The American Journal of Medicine* 72.2 (1982). Symposium on Disorders of Extracellular Volume and Composition: Part I, pp. 233 –240. ISSN: 0002-9343. DOI: http://dx.doi.org/10.1016/0002-9343(82)90815-4.

[305] Abu Z M Saleh et al. "Regulated proteolysis of the IFNaR2 subunit of the interferon-alpha receptor". In: *Oncogene* 23.42 (Aug. 2004), pp. 7076–7086. ISSN: 0950-9232. DOI: 10.1038/sj.onc.1207955.

[306] Diego Sanchez et al. "Aging without Apolipoprotein D: Molecular and cellular modifications in the hippocampus and cortex". In: *Experimental Gerontology* 67 (2015), pp. 19 –47. ISSN: 0531-5565. DOI: http://dx.doi.org/10.1016/j.exger.2015.04.003.

[307] Hakan Savli et al. "Gene network and canonical pathway analysis in prostate cancer: a microarray study". In: *Exp Mol Med* 40 (Apr. 2008), pp. 176–185. ISSN: 1226-3613. DOI: 10.3858/emm.2008.40.2.176.

[308] Uday Saxena. "Lipid metabolism and Alzheimer's disease: pathways and possibilities". In: *Expert Opinion on Therapeutic Targets* 13.3 (2009). PMID: 19236155, pp. 331–338. DOI: 10.1517/14728220902738720. eprint: http://dx.doi.org/10.1517/14728220902738720.

[309] C. D. Scharer et al. "Genome-wide promoter analysis of the SOX4 transcriptional network in prostate cancer cells". In: *Cancer research* 69.2 (2009), pp. 709–17. ISSN: 1538-7445 (Electronic) 0008-5472 (Linking). DOI: 10.1158/0008-5472.CAN-08-3415.

[310] Sophia Schedin-Weiss, Bengt Winblad and Lars O. Tjernberg. "The role of protein glycosylation in Alzheimer disease". In: *FEBS Journal* 281.1 (2014), pp. 46–62. ISSN: 1742-4658. DOI: 10.1111/febs.12590.

[311] Lena Scheubert et al. "Tissue-based Alzheimer gene expression markersвГY-comparison of multiple machine learning approaches and investigation of redundancy in small biomarker sets". English. In: *BMC Bioinformatics* 13.1, 266 (2012). DOI: 10.1186/1471-2105-13-266.

[312] Sara Schilit and Kenza E. Benzeroual. "Silodosin: A selective alpha1A-adrenergic receptor antagonist for the treatment of benign prostatic hyperplasia". In: *Clinical Therapeutics* 31.11 (2009), pp. 2489 –2502. ISSN: 0149-2918. DOI: 10.1016/j.clinthera.2009.11.024.

[313] N. Schuff et al. "MRI of hippocampal volume loss in early Alzheimer's disease in relation to ApoE genotype and biomarkers". In: *Brain* 132.4 (2009), pp. 1067–1077. ISSN: 0006-8950. DOI: 10.1093/brain/awp007.

[314] Wolfgang Schulz et al. "Factor interaction analysis for chromosome 8 and DNA methylation alterations highlights innate immune response suppression and cytoskeletal changes in prostate cancer". In: *Molecular Cancer* 6.1 (2007), p. 14. ISSN: 1476-4598.

[315] WolfgangA Schulz et al. "Changes in cortical cytoskeletal and extracellular matrix gene expression in prostate cancer are related to oncogenic ERG deregulation". In: *BMC Cancer* 10.1 (2010), pp. 1–9. DOI: 10.1186/1471-2407-10-505.

[316] Shobana Sekar et al. "Alzheimer's disease is associated with altered expression of genes involved in immune response and mitochondrial processes in astrocytes". In: *Neurobiology of Aging* 36.2 (Feb. 2015), pp. 583–591. ISSN: 0197-4580. DOI: 10.1016/j.neurobiolaging.2014.09.027.

[317] Lorenzo F Sempere et al. "Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation". In: *Genome Biology* 5.3 (Jan. 2004), R13–R13. ISSN: 1465-6914. DOI: 10.1186/gb-2004-5-3-r13.

[318] Stephanie Seneff, Glyn Wainwright and Luca Mascitelli. "Nutrition and Alzheimer's disease: The detrimental role of a high carbohydrate diet". In: *European Journal of Internal Medicine* 22.2 (2011), pp. 134 –140. ISSN: 0953-6205. DOI: http://dx.doi.org/10.1016/j.ejim.2010.12.017.

[319] Alberto Serrano-Pozo et al. "Neuropathological Alterations in Alzheimer Disease". In: *Cold Spring Harbor Perspectives in Medicine:* 1.1 (Sept. 2011), a006189–. ISSN: 2157-1422. DOI: 10.1101/cshperspect.a006189.

[320] Kaushik Shah, Shanal DeSilva and Thomas Abbruscato. "The Role of Glucose Transporters in Brain Disease: Diabetes and Alzheimer's Disease". In: *International Journal of Molecular Sciences* 13.10 (Sept. 2012), pp. 12629–12655. ISSN: 1422-0067. DOI: 10.3390/ijms131012629.

[321] Samuel Shapiro. "Is meta-analysis a valid approach to the evaluation of small effects in observational studies?" In: *Journal of Clinical Epidemiology* 50.3 (1997), pp. 223 –229. ISSN: 0895-4356. DOI: 10.1016/S0895-4356(96)00360-5.

[322] Masoud Shekarabi et al. "Mutations in the nervous system-specific HSN2 exon of WNK1 cause hereditary sensory neuropathy type II". In: *The Journal of Clinical Investigation* 118.7 (Apr. 2008), pp. 2496–2505. ISSN: 0021-9738. DOI: 10.1172/JCI34088.

[323] Hui Shen et al. "Critical and opposing roles of the chemokine receptors CXCR2 and CXCR3 in prostate tumor growth". In: *The Prostate* 66.16 (2006), pp. 1721–1728. ISSN: 1097-0045. DOI: 10.1002/pros.20476.

[324] Ng Shi-Yan et al. "Long noncoding RNA in development and disease of the central nervous system". In: *Trends in Genetics* 29.8 (2013), pp. 461 –468. ISSN: 0168-9525. DOI: http://dx.doi.org/10.1016/j.tig.2013.03.002.

[325] Dinesh Singh et al. "Gene expression correlates of clinical prostate cancer behavior". In: *Cancer Cell* 1 (2002), pp. 203–209–. ISSN: 1535-6108.

[326] Damian Smedley et al. "The BioMart community portal: an innovative alternative to large, centralized data repositories". In: *Nucleic Acids Research* 43 (2015), W589–W598. DOI: 10.1093/nar/gkv350. eprint: http://nar.oxfordjournals.org/content/early/2015/04/20/nar.gkv350.full.pdf+html.

[327] J Carson Smith et al. "Physical activity reduces hippocampal atrophy in elders at genetic risk for Alzheimer's disease". In: *Frontiers in Aging Neuroscience* 6 (Mar. 2014), pp. 61–. ISSN: 1663-4365. DOI: 10.3389/fnagi.2014.00061.

[328] Teresa C. Smith, David J. Spiegelhalter and Andrew Thomas. "Bayesian approaches to random-effects meta-analysis: A comparative study". In: *Statistics in Medicine* 14.24 (1995), pp. 2685–2699. ISSN: 1097-0258. DOI: 10.1002/sim.4780142408.

[329] A. Solodkin and G.W. Van Hoesen. "Entorhinal cortex modules of the human brain". In: *The Journal of Comparative Neurology* 365.4 (1996), pp. 610–627. ISSN: 1096-9861. DOI: 10.1002/(SICI)1096-9861(19960219)365:4<610::AID-CNE8>3.0.CO;2-7.

[330] Gabriela Spulber et al. "An MRI-based index to measure the severity of Alzheimer's disease-like structural pattern in subjects with mild cognitive impairment". In: *Journal of internal medicine* 273.4 (Jan. 2013), pp. 396–409. ISSN: 1365-2796. DOI: 10.1111/joim.12028.

[331] Jonathan Stallings et al. "Patterns of gene expression associated with recovery and injury in heat-stressed rats". In: *BMC Genomics* 15.1 (2014), p. 1058. ISSN: 1471-2164. DOI: 10.1186/1471-2164-15-1058.

[332] D. Stekel. *Microarray Bioinformatics*. Cambridge University Press, 2003. ISBN: 9780521525879.

[333] Yaakov Stern. "Cognitive reserve in ageing and Alzheimer's disease". In: *Lancet neurology* 11.11 (Nov. 2012), pp. 1006–1012. ISSN: 1474-4465. DOI: 10.1016/S1474-4422(12)70191-6.

[334] Mervyn Stone. "Cross-validatory choice and assessment of statistical predictions". In: *Journal of the Royal Statistical Society. Series B (Methodological)* (1974), pp. 111–147.

[335] Edward G. Stopa et al. "Basic fibroblast growth factor in Alzheimer's disease". In: *Biochemical and Biophysical Research Communications* 171.2 (1990), pp. 690–696. ISSN: 0006-291X. DOI: http://dx.doi.org/10.1016/0006-291X(90)91201-3.

[336] Samuel A. Stouffer et al. "The american soldier: Adjustment during army life. volume i". In: *Journal of the American Medical Association* 140.14 (1949), p. 1189. DOI: 10.1001/jama.1949.02900490055028. eprint: /data/Journals/JAMA/7618/jama_140_14_028.pdf.

[337] F. Strittmatter et al. "Inhibition of adrenergic human prostate smooth muscle contraction by the inhibitors of c-Jun N-terminal kinase, SP600125 and BI-78D3". In: *British Journal of Pharmacology* 166.6 (2012), pp. 1926–1935. ISSN: 1476-5381. DOI: 10.1111/j.1476-5381.2012.01919.x.

[338] Donna F Stroup et al. "Meta-analysis of observational studies in epidemiology: A proposal for reporting". In: *JAMA* 283.15 (2000), pp. 2008–2012. DOI: 10.1001/jama.283.15.2008. eprint: /data/Journals/JAMA/4732/JST00003.pdf.

[339] Donald T. Stuss, Catherine A. Gow and C. Ross Hetherington. ""No longer gage": Frontal lobe dysfunction and emotional changes." In: *Journal of Consulting and Clinical Psychology* 60.3 (1992), pp. 349–359. ISSN: 1939-2117(Electronic);0022006X(Print). DOI: 10.1037/0022-006X.60.3.349.

[340] Srinivasa Subramaniam et al. "ERK activation promotes neuronal degeneration predominantly through plasma membrane damage and independently of caspase-3". In: *The Journal of Cell Biology* 165.3 (2004), pp. 357–369. DOI: 10.1083/jcb.200403028. eprint: http://jcb.rupress.org/content/165/3/357.full.pdf+html.

[341] Patrick F. Sullivan, Michael C. Neale and Kenneth S. Kendler. "Genetic Epidemiology of Major Depression: Review and Meta-Analysis". In: *American Journal of Psychiatry* 157.10 (2000). PMID: 11007705, pp. 1552–1562. DOI: 10.1176/appi.ajp.157.10.1552. eprint: http://dx.doi.org/10.1176/appi.ajp.157.10.1552.

[342] Xutong Sun et al. "Down-regulation of WNK1 protein kinase in neural progenitor cells suppresses cell proliferation and migration". In: *Journal of Neurochemistry* 99.4 (2006), pp. 1114–1121. ISSN: 1471-4159. DOI: 10.1111/j.1471-4159.2006.04159.x.

[343] A.J. Sutton. *Methods for Meta-Analysis in Medical Research*. Wiley Series in Probability and Statistics - Applied Probability and Statistics Section. Wiley, 2000. ISBN: 9780471490661.

[344] Kirk Szafranski, Karan Joshua Abraham and Karim Mekhail. "Non-coding RNA in neural function, disease, and aging". In: *Frontiers in Genetics* 6.87 (2015). ISSN: 1664-8021. DOI: 10.3389/fgene.2015.00087.

[345] Damian Szklarczyk et al. "STRING v10: protein-protein interaction networks, integrated over the tree of life". In: *Nucleic Acids Research* 43.Database issue (Sept. 2014), pp. D447–D452. ISSN: 1362-4962. DOI: 10.1093/nar/gku1003.

[346] Michelle G. Tan et al. "Genome wide profiling of altered gene expression in the neocortex of Alzheimer's disease". In: *Journal of Neuroscience Research* 88.6 (2010), pp. 1157–1169. ISSN: 1097-4547. DOI: 10.1002/jnr.22290.

[347] Miyako Taniguchi and Katsuya Urakami. "Altered glycosylation in serum proteins of Alzheimer's disease". In: *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* 8.4 (2012), p. 119. DOI: 10.1016/j.jalz.2012.05.311.

[348] Juliet M. Taylor et al. "Type-1 interferon signaling mediates neuro-inflammatory events in models of Alzheimer's disease". In: *Neurobiology of Aging* 35.5 (2013), pp. 1012–1023. DOI: 10.1016/j.neurobiolaging.2013.10.089.

[349] R Development Core Team. "R: a language and environment for statistical computing". In: *Vienna, Austria : the R Foundation for Statistical Computing.* (2011).

[350] Hemlata Thackare, Helen D. Nicholson and Kate Whittington. "OxytocinвЃ"Yits role in male reproduction and new potential therapeutic uses". In: *Human Reproduction Update* 12.4 (2006), pp. 437–448. DOI: 10.1093/humupd/dmk002.

[351] Jeffrey G Thomas et al. "An Efficient and Robust Statistical Modeling Approach to Discover Differentially Expressed Genes Using Genomic Expression Profiles". In: *Genome Research* 11.7 (Apr. 2001), pp. 1227–1236. ISSN: 1088-9051. DOI: 10.1101/gr.165101.

[352] Simon G Thompson and Julian PT Higgins. "Can meta-analysis help target interventions at individuals most likely to benefit?" In: *The Lancet* 365.9456 (2005), pp. 341 –346. ISSN: 0140-6736. DOI: 10.1016/S0140-6736(05)17790-3.

[353] Mengnan Tian and Robert L. Macdonald. "The Intronic GABRG2 Mutation, IVS6+2T → G, Associated with Childhood Absence Epilepsy Altered Subunit mRNA Intron Splicing, Activated Nonsense-Mediated Decay, and Produced a Stable Truncated y2 Subunit". In: *The Journal of Neuroscience* 32.17 (2012), pp. 5937–5952. DOI: 10.1523/JNEUROSCI.5332-11.2012. eprint: http://www.jneurosci.org/content/32/17/5937.full.pdf+html.

[354] Leonard Henry Caleb Tippett. "The Methods of Statistics". In: *London: Williams and Norgate Ltd* (1931), pp. 7–284.

[355] Maria Christina Tsourlakis et al. "Overexpression of the chromatin remodeler death-domainвЃ"Yassociated protein in prostate cancer is an independent predictor of early prostate-specific antigen recurrence". In: *Human Pathology* 44.9 (2013), pp. 1789–1796. ISSN: 0046-8177. DOI: http://dx.doi.org/10.1016/j.humpath.2013.01.022.

[356] Sachiko Ueno et al. "An acidic fibroblast growth factor-like factor secreted into the brain cell culture medium upregulates apoE synthesis, {HDL} secretion and cholesterol metabolism in rat astrocytes". In: *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1589.3 (2002), pp. 261 –272. ISSN: 0167-4889. DOI: http://dx.doi.org/10.1016/S0167-4889(02)00181-7.

[357] Wullner Ulrich et al. "Cell-Specific Induction of Apoptosis by Rationally Designed Bivalent Aptamer-siRNA Transcripts Silencing Eukaryotic Elongation Factor 2". In: *Current Cancer Drug Targets* 8 (7) (2008), pp. 554–565. DOI: 10.2174/156800908786241078.

[358] Chandran Uma et al. "Gene expression profiles of prostate cancer reveal involvement of multiple molecular pathways in the metastatic process". In: *BMC Cancer* 7 (2007), pp. 64–. ISSN: 1471-2407.

[359] Sudarshan C Upadhya and Ashok N Hegde. "Role of the ubiquitin proteasome system in Alzheimer's disease". In: *BMC Biochemistry* 8.Suppl 1 (Nov. 2007), S12–S12. ISSN: 1471-2091. DOI: 10.1186/1471-2091-8-S1-S12.

[360] Roy Varshavsky et al. "Novel Unsupervised Feature Filtering of Biological Data". In: *Bioinformatics* 22.14 (2006), e507–e513. DOI: 10.1093/bioinformatics/btl214. eprint: http://bioinformatics.oxfordjournals.org/content/22/14/e507.full.pdf+html.

[361] Luciano de Gois Vasconcelos et al. "The thickness of posterior cortical areas is related to executive dysfunction in Alzheimer's disease." en. In: *Clinics* 69 (Jan. 2014), pp. 28 –37. ISSN: 1807-5932. DOI: 10.6061/clinics/2014(01)05.

[362] Jennifer Villasenor-Park and Alex G Ortega-Loayza. "Microarray Technique, Analysis, and Applications in Dermatology". In: *J Invest Dermatol* 133.4 (Apr. 2013), e7. ISSN: 0022-202X. DOI: 10.1038/jid.2013.64.

[363] Andrei G. Vlassenko, Tammie L. S. Benzinger and John C. Morris. "PET amyloid-beta imaging in preclinical Alzheimer's disease". In: *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1822.3 (2012), pp. 370–379. ISSN: 0925-4439. DOI: http://dx.doi.org/10.1016/j.bbadis.2011.11.005.

[364] P.C. Walsh and J.F. Worthington. *Dr. Patrick Walsh's Guide to Surviving Prostate Cancer*. Grand Central Publishing, 2013. ISBN: 9781455576173.

[365] J. Wang et al. "Strain- and region-specific gene expression profiles in mouse brain in response to chronic nicotine treatment". In: *Genes, Brain and Behavior* 7.1 (2008), pp. 78–87. ISSN: 1601-183X. DOI: 10.1111/j.1601-183X.2007.00328.x.

[366] Liangjiang Wang, Anand Srivastava and Charles Schwartz. "Microarray data integration for genome-wide analysis of human tissue-selective gene expression". In: *BMC Genomics* 11.Suppl 2 (2010), S15. ISSN: 1471-2164. DOI: 10.1186/1471-2164-11-S2-S15.

[367] Lin Wang et al. "ERG-SOX4 interaction promotes epithelial–mesenchymal transition in prostate cancer cells". In: *The Prostate* 74.6 (2014), pp. 647–658. ISSN: 1097-0045. DOI: 10.1002/pros.22783.

[368] Wenting Wang et al. "iBAG: integrative Bayesian analysis of high-dimensional multiplatform genomics data". In: *Bioinformatics* 29.2 (2013), pp. 149–159. DOI: 10.1093/bioinformatics/bts655.

[369] Xujing Wang, Soumitra Ghosh and Sun-Wei Guo. "Quantitative quality control in microarray image processing and data acquisition". In: *Nucleic Acids Research* 29.15 (June 2001), e75–e75. ISSN: 1362-4962.

[370] Yu Wang et al. "Gene selection from microarray data for cancer classification-a machine learning approach". In: *Computational Biology and Chemistry* 29.1 (2005), pp. 37 –46. ISSN: 1476-9271. DOI: 10.1016/j.compbiolchem.2004.11.001.

[371] G.F. Weber. *Molecular Mechanisms of Cancer*. SpringerLink: Springer e-Books. Springer, 2007. ISBN: 9781402060168.

[372] S. Wedel et al. "CXC chemokine mRNA expression as a potential diagnostic tool in prostate cancer". In: *Molecular Medicine Reports* 1.2 (2008), pp. 257–262.

[373] R. Weinberg. *The Biology of Cancer, Second Edition*. Taylor & Francis Group, 2013. ISBN: 9781317963462.

[374] John B. Welsh et al. "Analysis of Gene Expression Identifies Candidate Markers and Pharmacological Targets in Prostate Cancer". In: *Cancer Research* 61 (2001), pp. 5974–5978–.

[375] A. Whitehead. *Meta-Analysis of Controlled Clinical Trials*. Statistics in Practice. Wiley, 2002. ISBN: 9780471983705.

[376] Donna M. Wilcock. "Neuroinflammation in the Aging Down Syndrome Brain; Lessons from Alzheimer's Disease". In: *Current Gerontology and Geriatrics Research* 2012 (2012), p. 10. DOI: 10.1155/2012/170276.

[377] Donna M Wilcock and W Sue T Griffin. "Down's syndrome, neuroinflammation, and Alzheimer neuropathogenesis". In: *Journal of Neuroinflammation* 10 (May 2013), pp. 84–84. ISSN: 1742-2094. DOI: 10.1186/1742-2094-10-84.

[378] Claire L. Wilson and Crispin J. Miller. "Simpleaffy: a BioConductor package for Affymetrix Quality Control and data analysis". In: *Bioinformatics* 21.18 (2005), pp. 3683–3685. DOI: 10.1093/bioinformatics/bti605. eprint: http://bioinformatics.oxfordjournals.org/content/21/18/3683.full.pdf+html.

[379] Robert S. Wilson and David A. Bennett. "Cognitive Activity and Risk of Alzheimer's Disease". In: *Current Directions in Psychological Science* 12.3 (2003), pp. 87–91. DOI: 10.1111/1467-8721.01236. eprint: http://cdp.sagepub.com/content/12/3/87.full.pdf+html.

[380] Barnet Woolf. "Onestimating the relation between blood group and Disease." In: *Annals of Human Genetics* 19.4 (1955), pp. 251–253. ISSN: 1469-1809. DOI: 10.1111/j.1469-1809.1955.tb01348.x.

[381] Haitao Wu et al. "Sema4C Expression in Neural Stem/Progenitor Cells and in Adult Neurogenesis Induced by Cerebral Ischemia". English. In: *Journal of Molecular Neuroscience* 39.1-2 (2009), pp. 27–39. ISSN: 0895-8696. DOI: 10.1007/s12031-009-9177-8.

[382] Qian Wu, Rajiv Dhir and Alan Wells. "Altered CXCR3 isoform expression regulates prostate cancer cell migration and invasion". In: *Molecular Cancer* 11.1 (2012), p. 3. ISSN: 1476-4598.

[383] Ulrich Wullner et al. "Cell-specific induction of apoptosis by rationally designed bivalent aptamer-siRNA transcripts silencing eukaryotic elongation factor 2". In: *Current cancer drug targets* 8.7 (2008), pp. 554–565. ISSN: 1568-0096. DOI: 10.2174/156800908786241078.

[384] Charlie C. Xiang and Yidong Chen. "cDNA microarray technology and its applications". In: *Biotechnology Advances* 18.1 (2000), pp. 35 –46. ISSN: 0734-9750. DOI: http://dx.doi.org/10.1016/S0734-9750(99)00035-X.

[385] Fei Xiao et al. "Combined administration of D-galactose and aluminium induces Alzheimerlike lesions in brain". English. In: *Neuroscience Bulletin* 27.3 (2011), pp. 143–155. ISSN: 1673-7067. DOI: 10.1007/s12264-011-1028-2.

[386] Feifei Xiao et al. "miRecords: an integrated resource for microRNA-target interactions". In: *Nucleic Acids Research* 37.suppl 1 (2009), pp. D105–D110. DOI: 10.1093/nar/gkn851. eprint: http://nar.oxfordjournals.org/content/37/suppl_1/D105.full.pdf+html.

[387] Eric P Xing, Michael I Jordan, Richard M Karp et al. "Feature selection for high-dimensional genomic microarray data". In: *ICML*. Vol. 1. Citeseer. 2001, pp. 601–608.

[388] Bing-e Xu et al. "WNK1 Activates ERK5 by an MEKK2/3-dependent Mechanism". In: *Journal of Biological Chemistry* 279.9 (2004), pp. 7826–7831. DOI: 10.1074/jbc.M313465200. eprint: http://www.jbc.org/content/279/9/7826.full.pdf+html.

[389] Lei Xu, Donald Geman and Raimond Winslow. "Large-scale integration of cancer microarray data identifies a robust common cancer signature". In: *BMC Bioinformatics* 8.1 (2007), p. 275. ISSN: 1471-2105. DOI: 10.1186/1471-2105-8-275.

[390] Weili Xu et al. "Mid- and Late-Life Diabetes in Relation to the Risk of Dementia: A Population-Based Twin Study". In: *Diabetes* 58.1 (2009), pp. 71–77. DOI: 10.2337/db08-0586. eprint: http://diabetes.diabetesjournals.org/content/58/1/71.full.pdf+html.

[391] Y. Xu, H. Zhao and J. Hou. "Correlation between overexpression of Ep-CAM in prostate tissues and genesis of androgen-dependent prostate cancer". In: *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine* 35.7 (2014), pp. 6695–700. ISSN: 1423-0380 (Electronic) 1010-4283 (Linking). DOI: 10.1007/s13277-014-1892-2.

[392] Hidehisa Yamagata et al. "Promoter polymorphism in fibroblast growth factor 1 gene increases risk of definite Alzheimer's disease". In: *Biochemical and Biophysical Research Communications* 321.2 (2004), pp. 320 –323. ISSN: 0006-291X. DOI: http://dx.doi.org/10.1016/j.bbrc.2004.06.142.

[393] Kaori Yamamoto-Ishikawa et al. "The isolation and identification of apolipoprotein C-I in hormone-refractory prostate cancer using surface-enhanced laser desorption/ionization time-of-flight mass spectrometry". In: *Asian Journal of Andrology* 11.3 (Oct. 2008), pp. 299–307. ISSN: 1745-7262. DOI: 10.1038/aja.2008.38.

[394] Xinan Yang, Stefan Bentink and Rainer Spang. "Detecting Common Gene Expression Patterns in Multiple Cancer Outcome Entities". English. In: *Biomedical Microdevices* 7.3 (2005), pp. 247–251. ISSN: 1387-2176. DOI: 10.1007/s10544-005-3032-7.

[395] Xinan Yang and Xiao Sun. "Meta-analysis of several gene lists for distinct types of cancer: A simple way to reveal common prognostic markers". In: *BMC Bioinformatics* 8.1 (2007), p. 118. ISSN: 1471-2105. DOI: 10.1186/1471-2105-8-118.

[396] Eng-Juh Yeoh et al. "Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling". In: *Cancer Cell* 1.2 (2002), pp. 133 –143. ISSN: 1535-6108. DOI: 10.1016/S1535-6108(02)00032-6.

[397] Ka Yee Yeung, Roger E. Bumgarner and Adrian E. Raftery. "Bayesian model averaging: development of an improved multi-class, gene selection and classification tool for microarray data". In: *Bioinformatics* 21.10 (2005), pp. 2394–2402. DOI: 10.1093/bioinformatics/bti319. eprint: http://bioinformatics.oxfordjournals.org/content/21/10/2394.full.pdf+html.

[398] Aruto Yoshida et al. "Muscular Dystrophy and Neuronal Migration Disorder Caused by Mutations in a Glycosyltransferase, POMGnT1". In: *Developmental Cell* 1.5 (2001), pp. 717 –724. ISSN: 1534-5807. DOI: http://dx.doi.org/10.1016/S1534-5807(01)00070-3.

[399] Karen W. H. Young et al. "A Randomized, Crossover Trial of High-Carbohydrate Foods in Nursing Home Residents With Alzheimer's Disease: Associations Among Intervention Response, Body Mass Index, and Behavioral and Cognitive Function". In: *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 60.8 (2005), pp. 1039–1045. DOI: 10.1093/gerona/60.8.1039. eprint: http://biomedgerontology.oxfordjournals.org/content/60/8/1039.full.pdf+html.

[400] H-F Yuen et al. "Prostate cancer cells modulate osteoblast mineralisation and osteoclast differentiation through Id-1". In: *Br J Cancer* 102.2 (Dec. 2009), pp. 332–341. ISSN: 0007-0920. DOI: 10.1038/sj.bjc.6605480.

[401] C.C. Zai et al. "Association Study of GABRG2 Polymorphisms with Suicidal Behaviour in Schizophrenia Patients with Alcohol Use Disorder". In: *Neuropsychobiology* 69.3 (2014), pp. 154–158. ISSN: 0302-282X. DOI: 10.1159/000358839.

[402] Gwyneth Zai et al. "Evidence for the gamma-amino-butyric acid type B receptor 1 (GABBR1) gene as a susceptibility factor in obsessive-compulsive disorder". In: *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* 134B.1 (2005), pp. 25–29. ISSN: 1552-485X. DOI: 10.1002/ajmg.b.30152.

[403] Dmitri V Zaykin. "Optimally weighted Z-test is a powerful method for combining probabilities in meta-analysis". In: *Journal of evolutionary biology* 24.8 (May 2011), pp. 1836–1841. ISSN: 1420-9101. DOI: 10.1111/j.1420-9101.2011.02297.x.

[404] Ulrike Zeitschel et al. "Changes in activity and expression of phosphofructokinase in different rat brain regions after basal forebrain cholinergic lesion". In: *Journal of Neurochemistry* 83.2 (2002), pp. 371–380. ISSN: 1471-4159. DOI: 10.1046/j.1471-4159.2002.01127.x.

[405] Peng Zhang and Huaiyu Hu. "Differential glycosylation of alfadystroglycan and proteins other than alfa dystroglycan by like-glycosyltransfealfa". In: *Glycobiology* 22.2 (Aug. 2011), pp. 235–247. ISSN: 1460-2423. DOI: 10.1093/glycob/cwr131.

[406] Siqun L. Zheng et al. "Sequence Variants of Oç-Methylacyl-CoA Racemase Are Associated with Prostate Cancer Risk". In: *Cancer Research* 62.22 (2002), pp. 6485–6488.

[407] Miao Zhong et al. "Oxytocin Induces the Migration of Prostate Cancer Cells: Involvement of the Gi-Coupled Signaling Pathway". In: *Molecular Cancer Research* 8.8 (2010), pp. 1164–1172. DOI: 10.1158/1541-7786.MCR-09-0329. eprint: http://mcr.aacrjournals.org/content/8/8/1164.full.pdf+html.

[408] B. Zhu et al. "Tumor margin detection using quantitative NIRF molecular imaging targeting EpCAM validated by far red gene reporter iRFP". In: *Molecular imaging and biology : MIB : the official publication of the Academy of Molecular Imaging* 15.5 (2013), pp. 560–8. ISSN: 1860-2002 (Electronic) 1536-1632 (Linking). DOI: 10.1007/s11307-013-0637-8.

[409] X. Zhu et al. "The Role of Mitogen-Activated Protein Kinase Pathways in Alzheimer's Disease". In: *Neurosignals* 11.5 (2002), pp. 270–281. ISSN: 1424-862X. DOI: 10.1159/000067426.

[410] M N Ziats and O M Rennert. "Identification of differentially expressed microRNAs across the developing human brain". In: *Mol Psychiatry* 19.7 (July 2014), pp. 848–852. ISSN: 1359-4184. DOI: 10.1038/mp.2013.93.

[411] SamanthaM Zunich et al. "Osteoblast-secreted collagen upregulates paracrine Sonic hedgehog signaling by prostate cancer cells and enhances osteoblast differentiation". In: *Molecular Cancer* 11.1 (2012), pp. 1–13. DOI: 10.1186/1476-4598-11-30.