

# **Molecular Profiling of Lung Cancer**

Elizabeth Jane Richards

Imperial College London  
National Heart and Lung Institute  
Molecular Genetics and Genomics Group

Thesis of PhD  
2013

## **Statement of Originality**

The work described in this thesis was carried out in the laboratories of the Molecular Genetics Group, National Heart and Lung Institute (NHLI), Imperial College London. Unless otherwise stated, the author performed the experiments described. This thesis is submitted in support of my application for the degree of Doctor of Philosophy. It has been composed by the author and has not been submitted in any previous application for any degree. I confirm that it is an original work, representing my own academic effort and that all sources have been fully acknowledged.

Elizabeth Jane Richards

December 2013

## **Copyright Declaration**

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the license terms of this work.

## **Abstract**

Cisplatin is a first line chemotherapeutic agent for lung cancer however, although patients may respond to therapy, resistance often develops with tumour recurrence and disease progression. Somatic alterations in the tumour may alter therapeutic responses. Consequently a model of cisplatin resistance in lung cancer derived A549 cells has been created to examine the genomic changes that occur as chemo-resistance develops.

Drug resistance was induced in A549 cells through multiple rounds of cisplatin dosage and recovery over two different time courses. The concentration of cisplatin required to inhibit growth (inhibitory concentration [IC] value) was calculated at each round and cycles were continued until the IC value increased at least four-fold. Cells were harvested and total RNA extracted for whole transcriptome microarray analysis. Data was analysed using R statistics and associated packages, Affy, Limma, Mfuzz and WGCNA.

A five-fold increase in IC value was generated over successive doses in both regimes, accompanied by highly significant changes in gene expression. To explore these changes, temporal expression clustering and extensive network analyses were performed across the rounds of cisplatin dosing, as well as an untreated cell culture time course that acted as a comparison to the two treated regimes.

The results gathered from this robust model suggest that differences in dose and frequency of chemotherapy may affect genomic changes at specific loci that confer cisplatin resistance. Interesting and relevant pathways and genes have been discovered. In combination with analyses on a small patient cohort, these results have provided insights into the mechanism of cisplatin resistance and have highlighted new clinical biomarkers of potential use in prognosis of patients undergoing cancer treatment.

## Acknowledgements

I would like to thank my three supervisors, Miriam Moffatt, William Cookson, and Sanjay Popat, for all their input into my project. Without their expertise and guidance over the few years I would not be where I am today. Their support has been priceless and has helped me “push on” through all the inevitable upheavals. Miriam’s attention to detail, Bill’s focus on the big picture and Sanjay’s clinical input have come together to keep me on the straight and narrow. Many thanks go to the NHLI Foundation for sponsoring me through my time.

My colleagues and friends at the NHLI, both past and present, have helped keep me going through the best and worst times. The whole Molecular Genetics group have helped me grow from a naïve bright-eyed graduate to, hopefully, a more logical individual. Particular thanks must go to Saffron Willis-Owen whose patience with me in the field of bioinformatics have helped me to make sense of the data produced and assisted me in seeing where my future career lies. Further thanks must go to Mike Cox and Kenny Wong, who have put up with many a stupid question.

A more personal thank you goes to Frankie Bolt, Anna Thompson, Amy Turner, and Kirsty Russell for always being there, ready and waiting with a cup of tea when it was needed. Without their support, this work would not be complete.

Imperial College Boat Club must also be thanked for giving me a competitive distraction for the first couple of years of my PhD. The values of teamwork, leadership, determination and motivation have pushed me to strive for more, no matter what.

The final thank you has to go to my family in general and particularly to my mother. She has put up with every panic, every tear, and every rant over the past years and has been my rock. I could not have done it without you.

## Table of Contents

|                                                                                                                                                                       |           |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| STATEMENT OF ORIGINALITY                                                                                                                                              | 1         |
| COPYRIGHT DECLARATION                                                                                                                                                 | 1         |
| ABSTRACT                                                                                                                                                              | 2         |
| ACKNOWLEDGEMENTS                                                                                                                                                      | 3         |
| TABLE OF CONTENTS                                                                                                                                                     | 4         |
| LIST OF FIGURES                                                                                                                                                       | 8         |
| LIST OF TABLES                                                                                                                                                        | 11        |
| ABBREVIATIONS                                                                                                                                                         | 13        |
| <b>1 INTRODUCTION</b>                                                                                                                                                 | <b>16</b> |
| 1.1 Diagnosis and Symptoms                                                                                                                                            | 17        |
| 1.2 Types, Subtypes and Stages of Lung Cancer                                                                                                                         | 19        |
| 1.3 Causes of Lung Cancer                                                                                                                                             | 24        |
| 1.4 Lung Cancer Biology                                                                                                                                               | 25        |
| 1.5 Genetics, Genomics, Transcriptomics and Pharmacogenomics                                                                                                          | 29        |
| 1.5.1 EGFR ( <i>also known as HER1 or ERBB1</i> )                                                                                                                     | 30        |
| 1.5.2 KRAS                                                                                                                                                            | 34        |
| 1.5.3 ALK and Other Gene Rearrangements                                                                                                                               | 36        |
| 1.5.4 Developmental Genes: <i>PAX9, TTF1, NKX2-8</i> and <i>Shh</i>                                                                                                   | 38        |
| 1.5.5 DNA Repair and Nucleotide Excision Repair (NER): <i>ERCC1, RRM1 and BRCA1</i>                                                                                   | 39        |
| 1.5.6 Cell Cycle Regulators                                                                                                                                           | 43        |
| 1.5.7 Other Miscellaneous Somatic Mutations                                                                                                                           | 44        |
| 1.5.8 Other Alterations – <i>Structural Rearrangements, Copy Number Alterations, Mutant Allele Specific Imbalances, Uniparental Disomy and Loss of Heterozygosity</i> | 47        |
| 1.5.9 Germline genetics: Susceptibility Loci                                                                                                                          | 49        |
| 1.6 Treatment and Therapy                                                                                                                                             | 51        |
| 1.6.1 Molecular Mechanisms of Cisplatin Resistance                                                                                                                    | 53        |
| 1.7 Aims                                                                                                                                                              | 54        |
| <b>2 GENERAL MATERIALS AND METHODS</b>                                                                                                                                | <b>55</b> |

|            |                                                                                         |            |
|------------|-----------------------------------------------------------------------------------------|------------|
| <b>2.1</b> | <b>Materials</b>                                                                        | <b>55</b>  |
| 2.1.1      | Solutions and Reagents                                                                  | 55         |
| <b>2.2</b> | <b>Methods</b>                                                                          | <b>57</b>  |
| 2.2.1      | Cell Culture of A549 Cell Lines                                                         | 58         |
| 2.2.2      | Generation of Cisplatin-Resistant Cell Lines and Round Zero: IC25/IC50 Calculation      | 60         |
| 2.2.3      | Global Gene Expression                                                                  | 64         |
| 2.2.3.1    | <i>RNA Extraction from Cultured Cells</i>                                               | 64         |
| 2.2.3.2    | <i>RNA Preparation and Microarray Hybridisation for Global Gene Expression Analysis</i> | 66         |
| 2.2.3.3    | <i>Global Gene Expression Microarray Quality Control and Normalisation</i>              | 73         |
| 2.2.3.4    | <i>Differential Expression Analysis</i>                                                 | 73         |
| 2.2.3.5    | <i>Temporal Expression Profiling by Fuzzy C-means Clustering</i>                        | 74         |
| 2.2.3.6    | <i>Weighted Gene Co-expression Network Analysis</i>                                     | 75         |
| 2.2.4      | DNA Extraction from Cultured Cells                                                      | 75         |
| 2.2.5      | Lung Cancer Patient Tissue Work                                                         | 76         |
| 2.2.5.1    | <i>Collection of Patient Tissue and RNA Extraction</i>                                  | 76         |
| <b>3</b>   | <b>GENERATION AND OPTIMISATION OF A MODEL OF CISPLATIN RESISTANCE USING A549 CELLS</b>  | <b>79</b>  |
| <b>3.1</b> | <b>Introduction to Cell Culture Models</b>                                              | <b>79</b>  |
| <b>3.2</b> | <b>Optimisation of Repetitive Cisplatin Dosing and Associated Techniques</b>            | <b>82</b>  |
| 3.2.1      | Variability due to Seeding Density                                                      | 85         |
| 3.2.2      | Cisplatin Stability                                                                     | 89         |
| <b>3.3</b> | <b>Trials of Dosing using Optimised Conditions</b>                                      | <b>92</b>  |
| <b>3.4</b> | <b>Increasing Cisplatin Resistance</b>                                                  | <b>99</b>  |
| <b>3.5</b> | <b>Chemo-naïve Normal Cell Time Course</b>                                              | <b>101</b> |
| <b>3.6</b> | <b>Discussion</b>                                                                       | <b>101</b> |
| <b>4</b>   | <b>GLOBAL GENE EXPRESSION ANALYSIS OF CISPLATIN RESISTANCE</b>                          | <b>106</b> |
| <b>4.1</b> | <b>Introduction to Global Gene Expression Analysis</b>                                  | <b>106</b> |
| <b>4.2</b> | <b>RNA Preparation: Extraction and Quality Control</b>                                  | <b>108</b> |
| 4.2.1      | RNA Extraction from Cultured Cells                                                      | 108        |
| 4.2.2      | Microarray Quality Control, Normalisation and Transformation                            | 110        |
| 4.2.2.1    | <i>IC50 Dosing 3 Week Recovery</i>                                                      | 111        |
| 4.2.2.2    | <i>IC25 Dosing 1 Week Recovery</i>                                                      | 115        |
| <b>4.3</b> | <b>Differential Gene Expression Analysis</b>                                            | <b>117</b> |
| 4.3.1      | Top Differentially Expressed Genes from the Two Treated Time Courses                    | 118        |
| <b>4.4</b> | <b>Temporal Cluster Analysis</b>                                                        | <b>123</b> |

|            |                                                                                                                                              |            |
|------------|----------------------------------------------------------------------------------------------------------------------------------------------|------------|
| 4.4.1      | Expression Clustering                                                                                                                        | 125        |
| 4.4.1.1    | <i>IC50 Dose 3 Week Recovery Time Course</i>                                                                                                 | 125        |
| 4.4.1.2    | <i>IC25 Dose 1 Week Recovery Time Course</i>                                                                                                 | 128        |
| <b>4.5</b> | <b>Gene Expression Analysis of Untreated A549 Cells</b>                                                                                      | <b>131</b> |
| 4.5.1      | Comparisons of the Significantly Differentially Expressed Genes between the Untreated and Treated Time Courses                               | 132        |
| 4.5.2      | Investigating the Overlap between the Time Courses                                                                                           | 138        |
| <b>4.6</b> | <b>Discussion</b>                                                                                                                            | <b>143</b> |
| <b>5</b>   | <b>NETWORK ANALYSIS</b>                                                                                                                      | <b>149</b> |
| <b>5.1</b> | <b>Introduction to Network Analysis</b>                                                                                                      | <b>149</b> |
| <b>5.2</b> | <b>Weighted Gene Co-expression Network Analysis</b>                                                                                          | <b>150</b> |
| 5.2.1      | Creating the Network                                                                                                                         | 150        |
| 5.2.2      | Relating Modules with Traits of Interest                                                                                                     | 153        |
| 5.2.3      | IC50 Dose 3 Week Recovery Time Course Modules of Interest                                                                                    | 155        |
| 5.2.4      | IC25 Dose 1 Week Recovery Time Course Modules of Interest                                                                                    | 159        |
| 5.2.5      | Normal Untreated A549 Cells Time Course                                                                                                      | 164        |
| 5.2.6      | Differentially Expressed Genes in both Treated Time Courses but not present in the Untreated Time Course                                     | 165        |
| <b>5.3</b> | <b>WGCNA Discussion</b>                                                                                                                      | <b>168</b> |
| <b>6</b>   | <b>GENE EXPRESSION ANALYSIS OF PATIENT SAMPLES</b>                                                                                           | <b>177</b> |
| <b>6.1</b> | <b>The Need for Validation in Patient Samples</b>                                                                                            | <b>177</b> |
| <b>6.2</b> | <b>RESTORE-AIR – Randomised Evaluation of STents to Open REstricted AIRways in patients with centrally placed non-small cell lung cancer</b> | <b>179</b> |
| <b>6.3</b> | <b>EQUALITY – Expression QUAntitative trait Loci mapping In Tumours collected at biopsY</b>                                                  | <b>180</b> |
| <b>6.4</b> | <b>RNA Extraction from Tissue Samples</b>                                                                                                    | <b>181</b> |
| <b>6.5</b> | <b>A Dataset for Comparisons of Platinum Treatment versus Non-Treated Samples</b>                                                            | <b>183</b> |
| 6.5.1      | Outlier Detection and Removal, and Microarray Quality Control                                                                                | 184        |
| <b>6.6</b> | <b>Differential Expression Analysis</b>                                                                                                      | <b>190</b> |
| <b>6.7</b> | <b>Discussion</b>                                                                                                                            | <b>205</b> |
| <b>7</b>   | <b>DISCUSSION AND FUTURE WORK</b>                                                                                                            | <b>208</b> |
| <b>7.1</b> | <b>Discussion</b>                                                                                                                            | <b>208</b> |
| 7.1.1      | A549 Cellular Models of Cisplatin Resistance                                                                                                 | 208        |

|            |                                                                                                                       |            |
|------------|-----------------------------------------------------------------------------------------------------------------------|------------|
| 7.1.2      | Temporal Gene Expression Profiling                                                                                    | 209        |
| 7.1.3      | Network Analysis                                                                                                      | 211        |
| 7.1.4      | Potential Future Work: Cellular Model                                                                                 | 213        |
| 7.1.4.1    | <i>Other Cellular Models</i>                                                                                          | 213        |
| 7.1.4.2    | <i>Gene Expression over the Time Course</i>                                                                           | 213        |
| 7.1.4.3    | <i>Genome of the Cisplatin Resistance Model</i>                                                                       | 214        |
| 7.1.5      | Investigating Genes of Interest from the Cellular Model of Resistance between Platinum Treated and Untreated Patients | 215        |
| 7.1.5.1    | <i>Future Work: Tumour Tissue Samples</i>                                                                             | 216        |
| <b>7.2</b> | <b>General Conclusion</b>                                                                                             | <b>218</b> |
| <b>8</b>   | <b>BIBLIOGRAPHY</b>                                                                                                   | <b>221</b> |
| <b>9</b>   | <b>APPENDICES</b>                                                                                                     | <b>236</b> |
| <b>9.1</b> | <b>E-Appendices for use in conjunction with Chapter 4</b>                                                             | <b>236</b> |
| <b>9.2</b> | <b>E-Appendices for use in conjunction with Chapter 5</b>                                                             | <b>236</b> |
| <b>9.3</b> | <b>Circos Plots of Gene Interconnectivity within Modules</b>                                                          | <b>237</b> |
| <b>9.4</b> | <b>Copyright Permissions</b>                                                                                          | <b>246</b> |

## List of Figures

**Figure 1.1** Diagram showing a build-up of fluid between the visceral and parietal pleura of the lung.

**Figure 1.2** 40x magnification of Squamous Cell Carcinoma of the lung.

**Figure 1.3** 40x magnification of Adenocarcinoma of the lung.

**Figure 1.4** 200x magnification of Small Cell Carcinoma of the lung.

**Figure 1.5** Diagram illustrating the evolution of lung cancer that is smoking-related and non-smoking-related.

**Figure 1.6** Pathway diagram showing multiple pathways and their signalling molecules.

**Figure 1.7** Schematic showing dimerisation of EGFR, its signalling pathways and their subsequent influence on downstream cellular functions.

**Figure 1.8** Nucleotide Excision Repair.

**Figure 1.9** Three ways to acquire Mutant Allele Specific Imbalances (MASI).

**Figure 1.10** Cisplatin cross-links DNA.

**Figure 2.1** Diagram of the haemocytometer layout under a light microscope.

**Figure 2.2** Flow diagram outlining the repeated dosing of A549 cells with cisplatin.

**Figure 2.3** Example of an RNA electropherogram.

**Figure 2.4** Outline of the RNA processing protocol.

**Figure 2.5** Example of a cRNA electropherogram.

**Figure 2.6** Example of a cDNA electropherogram.

**Figure 2.7** Example of a fragmented cDNA electropherogram.

**Figure 3.1** Dose response with cisplatin in 0.9% NaCl solution after a 3 hour incubation for the concentration range 0.1  $\mu$ M to 300  $\mu$ M.

**Figure 3.2** Dose response with cis.platin in serum free media after a 3 hour incubation for the concentration range 0.1  $\mu$ M to 300  $\mu$ M

**Figure 3.3** Dose response of cells seeded at  $0.1 \times 10^6$  cells per well with triplicate cisplatin concentrations in serum free media incubated for 24 hours.

**Figure 3.4** Dose response using optimised conditions of 2,500 cells per well, 72 hour incubation with cisplatin in complete media with MTT performed immediately after drug removal.

**Figure 3.5** Final 96-well plate map of layout of concentrations for cisplatin dosing across two plates.

**Figure 3.6** IC<sub>50</sub> calculation showing drug degradation leading to apparently increased Round 0 IC<sub>50</sub> value compared to that calculated from multiple experiments with freshly prepared cisplatin.

**Figure 3.7** IC<sub>50</sub> calculation showing Round 0 IC<sub>50</sub> = 5  $\mu$ M (n=6, with SEM error bars).

**Figure 3.8** Normal cells 24 hours after split (seeding density of  $2 \times 10^6$  cells) showing healthy growth, with fairly uniform shape, size and structure (10x magnification).

**Figure 3.9** Cells at ten days recovery showing stressed characteristics, including elongation, loss of defined structure and granularisation.

**Figure 3.10** Cells at ten days into recovery showing both stressed and normal areas of cellular growth.

**Figure 3.11** IC50 dose with  $0.75 \times 10^6$  seeding density and two week recovery Round 1 IC50 dose calculation.

**Figure 3.12** IC25 dose with  $0.75 \times 10^6$  seeding density and one week recovery Round 3 IC25 dose calculation.

**Figure 3.13** Comparison of  $0.75 \times 10^6$  cell seed and  $2 \times 10^6$  cell seed over five rounds of IC50 dosing.

**Figure 3.14** Comparison of 2 week, 3 week and 4 week recovery periods over four to five rounds of IC50 dosing.

**Figure 3.15** Plot of the mean for IC25 values from the cisplatin dose response at each time point from chemo-naïve cells round 0 to round 15.

**Figure 3.16** Plot of the mean for IC50 values from the cisplatin dose response at each time point from chemo-naïve cells round 0 to round 10.

**Figure 4.1** An example of a false-coloured hybridised array as scanned using the GeneTitan.

**Figure 4.2** Log Expression Signal showing the box and whisker plots for each array file in the IC50 dose 3 week recovery time course.

**Figure 4.3** Signal Distribution Histogram for all 36 arrays from the IC50 dose 3 week recovery time course.

**Figure 4.4** Bacterial spike signal values generated in Expression Console, showing each array from the IC50 dose 3 week recovery time course and its intensity for each bacterial spike hybridisation control.

**Figure 4.5 Left:** Log Expression Signal of IC25 dose 1 week recovery. **Right:** Log Expression Signal of IC25 dose 1 week recovery on replacement of the Round 1 Replicate.

**Figure 4.6** Hierarchical clustering plot of expression data for all time points for the IC25 dose 1 week recovery time course.

**Figure 4.7** IC50 Dosing with 3 Week Recovery: fuzzy c-means cluster analysis.

**Figure 4.8** IC25 Dose with 1 Week Recovery: fuzzy c-means cluster analysis.

**Figure 4.9** Log Expression Signal Box and Whisker plot of Normal Time Course.

**Figure 4.10** Signal Density Distribution showing almost identical distributions across all the 33 arrays.

**Figure 4.11** Untreated Normal Time Course: fuzzy c-means cluster analysis.

**Figure 4.12** Venn diagram showing the overlap of 843 gene between the those that are most significantly differentially expressed in the IC50 dose 3 week recovery regime and the IC25 dose 1 week recovery regime.

**Figure 5.1** A scale-free topology plot for the IC50 dose 3 week recovery time course.

**Figure 5.2** The hierarchical cluster dendrogram showing the assignment of genes to modules (colour band) based on the TOM for the IC50 dose 3 week recovery time course.

**Figure 5.3** A module-trait relationship matrix for the IC50 dose 3 week recovery time course modules showing strength of correlation with the trait of interest (IC50 value).

**Figure 5.4** Circos plot of the Blue module from the IC50 dose 3 week recovery time course.

**Figure 5.5** A module-trait relationship matrix for the IC25 dose 1 week recovery time course modules showing strength of correlation with the trait of interest (IC25 value).

**Figure 5.6** A module-trait relationship matrix for the IC25 dose 1 week recovery time course modules, generated by cutting higher on the clustering dendrogram, reducing the number of modules, showing strength of correlation with the trait of interest (IC25 value).

**Figure 5.7** A module-trait relationship matrix for the 375 genes only differentially expressed in the treated time courses using the IC50 dose 3 week recovery time course data.

**Figure 6.1** Box and whisker plot of normalised intensities (log expression signal) of the seventeen patient samples.

**Figure 6.2** Hierarchical cluster dendrogram of all seventeen .CEL files using 1-IAC as the distance metric.

**Figure 6.3** Plot showing the distribution of the mean IAC for each array.

**Figure 6.4** Plot showing the distribution of mean IAC for each array after outlier removal with no .CEL files falling outside the standard deviation cut off.

**Figure 6.5** Principle Components Analysis plot showing the distribution of the three studies for the patient tissue samples.

**Figure 6.6** Principle Components Analysis plot showing how the batch effect of study has been removed to give a more even dataset within the patient tissue samples.

**Figure 6.7** Box and whisker plot showing the expression of *INA* across the samples, separated into non-platinum-treated and platinum-treated.

**Figure 6.8** Box and whisker plot showing the expression of *RAB15* across the samples, separated into non-platinum-treated and platinum-treated.

**Figure 6.9** Box and whisker plot showing the expression of *ABHD4* across the samples, separated into non-platinum-treated and platinum-treated.

**Figure 6.10** Box and whisker plot showing the expression of *AFAP1* across the samples, separated into non-platinum-treated and platinum-treated.

## List of Tables

**Table 1.1** Tumour/Node/Metastasis Staging in Lung Cancer.

**Table 2.1** Volumes required to process Gene 1.1 ST Array plates per run.

**Table 3.1** Table indicating the known and verified cancer point mutations in A549 cells.

**Table 3.2** Details of mean IC50 and IC25 values calculated for each round of cisplatin dosing under all conditions tested.

**Table 4.1** IC50 dose 3 week recovery time course; top 25 differentially expressed genes.

**Table 4.2** IC25 dose 1 week recovery time course; top 25 differentially expressed genes.

**Table 4.3** Number of genes within each expression cluster for the IC50 dose 3 week recovery time course.

**Table 4.4** Number of genes within each expression cluster for the IC25 dose 1 week recovery time course.

**Table 4.5** Normal untreated time course; top 25 differentially expressed genes.

**Table 4.6** Number of genes within each temporal expression cluster for the 865 differentially expressed genes from the untreated normal time course.

**Table 4.7** Comparison of the number of differentially expressed genes within each increasing or decreasing expression cluster with the up-regulated and down-regulated gene clusters in the untreated normal time course.

**Table 4.8** 68 unique genes that share similar increasing and decreasing expression profiles over the course of cisplatin treatment and extended untreated cell culture rounds.

**Table 4.9** 28 differentially expressed genes that overlap the two treated time courses with at least a three-fold change in expression between the start and end of the IC50 dose 3 week recovery time course.

**Table 4.10** Genes overlapping the two increasing expression profile clusters and the two decreasing expression profile clusters.

**Table 5.1** Most highly connected genes within modules that have the most significant associations with IC50 value.

**Table 5.2** Number of genes showing increasing and decreasing expression from the temporal expression clustering across modules of greatest correlation with IC50 value.

**Table 5.3** Number of genes showing increasing and decreasing expression from the temporal expression clustering across modules of greatest correlation with IC25 value.

**Table 5.4** Most highly connected genes within the reduced modules that have the most significant associations with IC25 value.

**Table 5.5** Number of genes showing increasing and decreasing expression from the temporal expression clustering, across the modules of greatest correlation with IC25 value for the reduced WGCNA modules.

**Table 5.6** Number of genes within each module colour for the Normal Untreated Time Course.

**Table 5.7** Highly enriched Gene Ontology functional annotation clusters for the modules created by WGCNA for the differentially expressed genes from the Normal Untreated Time Course.

**Table 5.8** Most highly connected genes from the 375 overlapping genes within modules that have the most significant association with IC50 value.

**Table 6.1** Demographics for the seven RESTORE-AIR participants.

**Table 6.2** Demographics for the EQUALITY patients.

**Table 6.3** RNA extraction from tissue for RESTORE-AIR and EQUALITY samples showing the resulting concentrations.

**Table 6.4** Demographics for Patient Samples for differential expression analyses.

**Table 6.5** Time course expression clusters and WGCNA modules used to perform differential expression analyses between treated and untreated patient tissues.

**Table 6.6** Top 21 genes, seen in patients, from differential expression analyses using gene lists of interest taken from the *in vitro* cellular model.

## Abbreviations

|                 |                                                                                  |
|-----------------|----------------------------------------------------------------------------------|
| AKT             | v-AKT Murine Thymoma Viral Oncogene Homolog 1                                    |
| ALK             | Anaplastic Lymphoma Kinase                                                       |
| APE1            | Apurinic/apyrimidinic Endonuclease 1                                             |
| APT             | Affymetrix Power Tools                                                           |
| ATCC            | American Type Culture Collection                                                 |
| ATP             | Adenosine Triphosphate                                                           |
| ANOVA           | Analysis of Variance                                                             |
| AUC             | Area Under Curve                                                                 |
| BCD             | Bisulfite Converted DNA                                                          |
| B-H             | Benjamini & Hochberg                                                             |
| BRAF            | v-RAF Murine Sarcoma Viral Oncogene Homolog B1                                   |
| CDK             | Cyclin-Dependent Kinases                                                         |
| cDNA            | Complementary Deoxyribonucleic Acid                                              |
| CNAs            | Copy Number Alterations                                                          |
| CNGs            | Copy Number Gains                                                                |
| CO <sub>2</sub> | Carbon Dioxide                                                                   |
| COPD            | Chronic Obstructive Pulmonary Disease                                            |
| CpG             | Cytosine-Phosphate-Guanine                                                       |
| cRNA            | Complementary Ribonucleic Acid                                                   |
| CRUK            | Cancer Research UK                                                               |
| DAVID           | Database for Annotation, Visualisation and Integrated Discovery                  |
| ddNTP           | Dideoxynucleotide                                                                |
| DEPC            | Diethylpyrocarbonate                                                             |
| DMEM            | Dulbecco's Modified Eagles Medium                                                |
| DMSO            | Dimethylsulfoxide                                                                |
| DNA             | Deoxyribonucleic Acid                                                            |
| EDTA            | Ethylenediaminetetraacetic Acid                                                  |
| EGF             | Epidermal Growth Factor                                                          |
| EGFR            | Epidermal Growth Factor Receptor                                                 |
| EML4            | Echinoderm Microtubule-Associated Protein-Like 4                                 |
| EMT             | Epithelial to Mesenchymal Transition                                             |
| eQTL            | Expression Quantitative Trait Loci                                               |
| EQUALITY        | Expression QUAntitative trait Loci mapping In Tumours collected at bronchoscopy  |
| ERBB            | Erythroblastic Leukemia Viral Oncogene Homolog                                   |
| ERCC            | Excision Repair Cross-Complementation                                            |
| ERK             | Extracellular Signal-Regulated Kinase/Mitogen-activated protein kinase 3 (MAPK3) |
| FCS             | Foetal Calf Serum                                                                |
| FDR             | False Discovery Rate                                                             |
| FFPE            | Formalin-Fixed Paraffin-Embedded                                                 |
| gDNA            | Genomic Deoxyribonucleic Acid                                                    |
| GO              | Gene Ontology                                                                    |
| GRI             | Genetic Research Instrumentation                                                 |
| GS              | Gene Significance                                                                |
| GWAS            | Genome Wide Association Study                                                    |

|             |                                                                |
|-------------|----------------------------------------------------------------|
| HBSS        | Hanks Balanced Salt Solution                                   |
| HER         | Human Epidermal Growth Factor Receptor                         |
| HLA         | Human Leukocyte Antigen                                        |
| HUGO        | Human Genome Organisation                                      |
| IAC         | Inter-Array Correlation                                        |
| IC25        | Inhibitory concentration required to inhibit growth by 25%     |
| IC50        | Inhibitory concentration required to inhibit growth by 50%     |
| IPA         | Ingenuity Pathway Analysis                                     |
| ISD         | Information Services Division                                  |
| IVT         | <i>In Vitro</i> Transcription                                  |
| kME         | Eigengene-Based Connectivity                                   |
| KRAS        | Kirsten Rat Sarcoma Viral Oncogene Homolog                     |
| Limma       | Linear Models for Microarray Data                              |
| LOH         | Loss of Heterozygosity                                         |
| MASI        | Mutant Allele Specific Imbalances                              |
| ME          | Module Eigengene                                               |
| MEK         | Mitogen-activated protein kinase kinase (MAP2K)                |
| MET         | Mesenchymal Epithelial Transcription Factor                    |
| miRNA       | Micro RNA                                                      |
| MM          | Module Membership                                              |
| MMP         | Matrix Metalloproteinase Family                                |
| MMR         | Mismatch Repair                                                |
| MTT         | 3-(4, 5-Dimethylthiazol-2-yl)-2, 5-diphenyltetrazolium bromide |
| NaCl        | Sodium Chloride                                                |
| NaOH        | Sodium Hydroxide                                               |
| NCBI        | National Centre for Biotechnology Information                  |
| NER         | Nucleotide Excision Repair                                     |
| NHEJ        | Non-Homologous End Joining                                     |
| NHLI        | National Heart and Lung Institute                              |
| NK-Kb       | Nuclear Factor kappa B                                         |
| NSCLC       | Non-Small Cell Lung Cancer                                     |
| OMIM        | Online Mendelian Inheritance in Man                            |
| ONS         | Office for National Statistics                                 |
| PAM         | Partitioning Around Medoids                                    |
| PBS         | Phosphate Buffered Saline                                      |
| PCA         | Principle Components Analysis                                  |
| PCR         | Polymerase Chain Reaction                                      |
| PI3K        | Phosphoinositide 3-Kinase                                      |
| PIS         | Patient Information Sheet                                      |
| PLIER       | Probe Logarithmic Intensity Error                              |
| PTEN        | Phosphatase and Tensin Homolog                                 |
| QC          | Quality Control                                                |
| qRT-PCR     | Quantitative Reverse Transcriptase PCR                         |
| REC         | Research Ethics Committee                                      |
| RESTORE-AIR | Randomised Evaluation of STents to Open REstricted AIRways     |
| RIN         | RNA Integrity Number                                           |
| RLE         | Relative Log Expression                                        |
| RMA         | Robust Multiarray Averaging                                    |
| RNA         | Ribonucleic Acid                                               |

|         |                                                                      |
|---------|----------------------------------------------------------------------|
| RNA-seq | RNA Sequencing                                                       |
| SCLC    | Small Cell Lung Cancer                                               |
| SD      | Standard Deviation                                                   |
| SEM     | Standard Error of the Mean                                           |
| Sf      | Significant Figures                                                  |
| SFM     | Serum Free Media                                                     |
| SGH     | St George's Hospital                                                 |
| SIADH   | Syndrome of Inappropriate Antidiuretic Hormone                       |
| SNP     | Single Nucleotide Polymorphism                                       |
| snRNA   | Small Nuclear RNA                                                    |
| ST      | Sense Target                                                         |
| TdT     | Terminal Deoxynucleotidyl Transferase                                |
| TKI     | Tyrosine Kinase Inhibitor                                            |
| TNM     | TNM Classification of Malignant Tumours<br>(Tumour/Nodes/Metastasis) |
| TNF     | Tumour Necrosis Factor                                               |
| TOM     | Topological Overlap Matrix                                           |
| TSG     | Tumour Suppressor Gene                                               |
| UDG     | Uracil-DNA glycosylase                                               |
| UICC    | Union for International Cancer Control                               |
| UK      | United Kingdom                                                       |
| UPD     | Uniparental Disomy                                                   |
| USA     | United States of America                                             |
| UV      | Ultra Violet                                                         |
| WGCNA   | Weighted Gene Co-expression Network Analysis                         |
| wt      | Wild Type                                                            |
| WT      | Whole Transcript                                                     |

# 1 Introduction

One in eight deaths worldwide is caused by cancer. There are more than 100 distinct tumour types, with diverse risk factors and epidemiology, which can originate from most of the cell types and organs of the human body. Lung cancer is the most common cause of cancer-related death globally and in the UK alone has an incidence rate of approximately 41,500 per annum (2009) and a mortality rate of approximately 35,000 per annum (2010), (data combined from the Office for National Statistics [ONS] 2011, Information Services Division [ISD] Scotland 2011, Welsh Cancer Intelligence and Surveillance Unit 2011, Northern Ireland Cancer Registry 2011). Worldwide 1.61 million people were diagnosed with lung cancer in 2008 (World Cancer Research Fund, <http://www.wcrf.org>) whilst in the UK it is the second most common cancer. By sex, lung cancer is the second most common in men, following prostate and the third most common in women after breast and bowel cancers (from data combined from ONS, ISD Scotland 2011, Welsh Cancer Intelligence and Surveillance Unit 2011, Northern Ireland Cancer Registry 2011). The majority (75%) of lung cancer cases occur in the over 65 years of age group (data combined from ONS 2007-2009, ISD Scotland 2007-2009, Welsh Cancer Intelligence and Surveillance Unit 2007-2009, Northern Ireland Cancer Registry 2007-2009) and most patients (68%) present with advanced disease, Stage III or IV (see Section 1.2), (National Cancer Registration Service/<http://www.cancerresearchuk.org>). There is a strong correlation between one-year relative survival rate and the stage of disease at diagnosis, i.e. those with later stage disease at diagnosis have a lower relative survival chance. This late stage presentation combined with relative chemo-refractoriness results in poor survival rates – less than 9% (not age-standardised) survive five years after diagnosis (ONS and ISD Scotland).

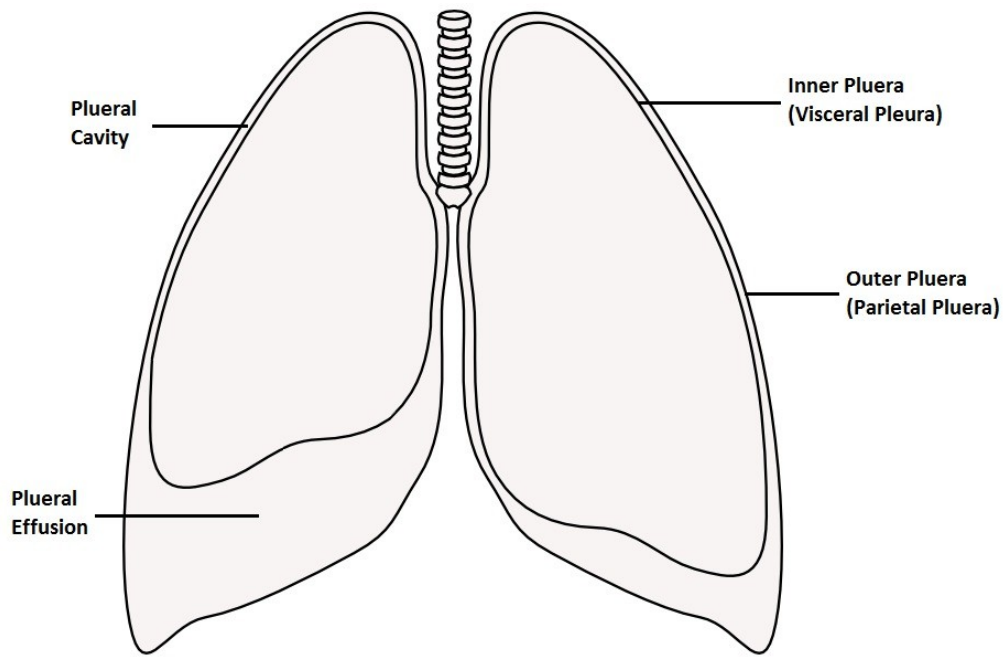
## 1.1 Diagnosis and Symptoms

Initial symptoms of lung cancer include:-

- Persistent coughing
- A change in a long term cough
- Shortness of breath
- Coughing up sputum with signs of blood in it (haemoptysis)
- An ache or pain when breathing or coughing (pleuritic pain)
- Loss of appetite
- Fatigue
- Weight loss

Less common symptoms, although they are generally indicative of more advanced or later stage lung cancer, include:-

- A hoarse voice
- Difficulty swallowing
- Finger clubbing
- Swelling of the face caused by an obstruction of the superior vena cava
- Swelling in the neck caused by enlarged lymph nodes
- Pain or discomfort under the ribs, from the liver
- Shortness of breath caused by pleural effusion (Figure 1.1)



**Figure 1.1** Diagram showing a build-up of fluid between the visceral and parietal pleura of the lungs (pleural effusion) which results in shortness of breath (<http://www.cancerresearchuk.org>).

Endocrine symptoms can occasionally present. Some types of lung cancer cells produce peptides or hormones that either enter the bloodstream or act on local organs and may cause symptoms that seem to be unrelated to the lung cancer. These 'paraneoplastic symptoms' are uncommon.

Symptoms vary from person to person but may include:

- Confusion or forgetfulness
- Pins and needles or numbness in the fingers or toes
- Muscle weakness
- General drowsiness, weakness, dizziness or confusion
- Thrombosis

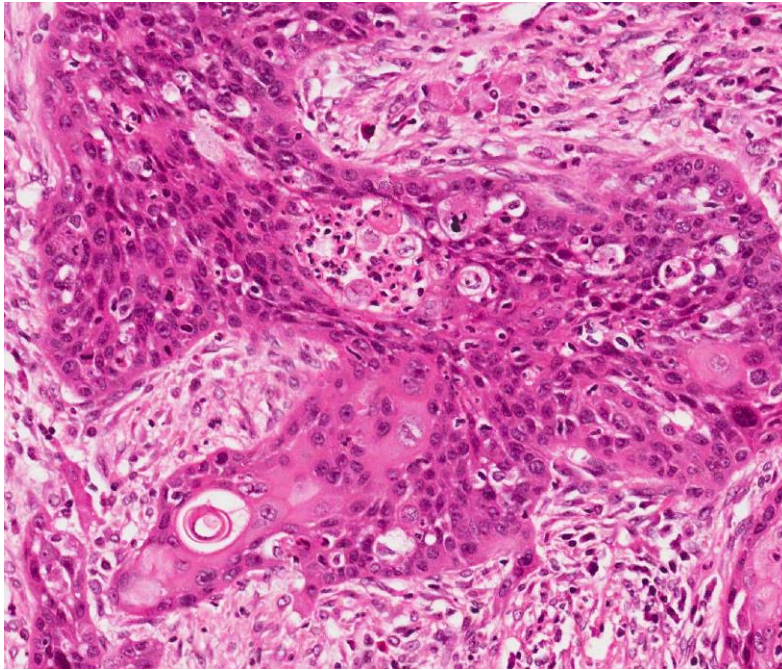
Lung cancers growing at the very top of the lung (apices) are called 'pancoast tumours'. These tumours can cause very specific symptoms, most commonly severe shoulder pain. They can

also cause 'Horner's syndrome' which is a collection of signs and symptoms caused by the tumour compressing the cervical sympathetic nerve chain. Features include drooping or weakness of one eyelid (ptosis), small pupil in the same eye (miosis), and loss of ability to sweat (anhidrosis) on the same side of the face in which the eye is affected.

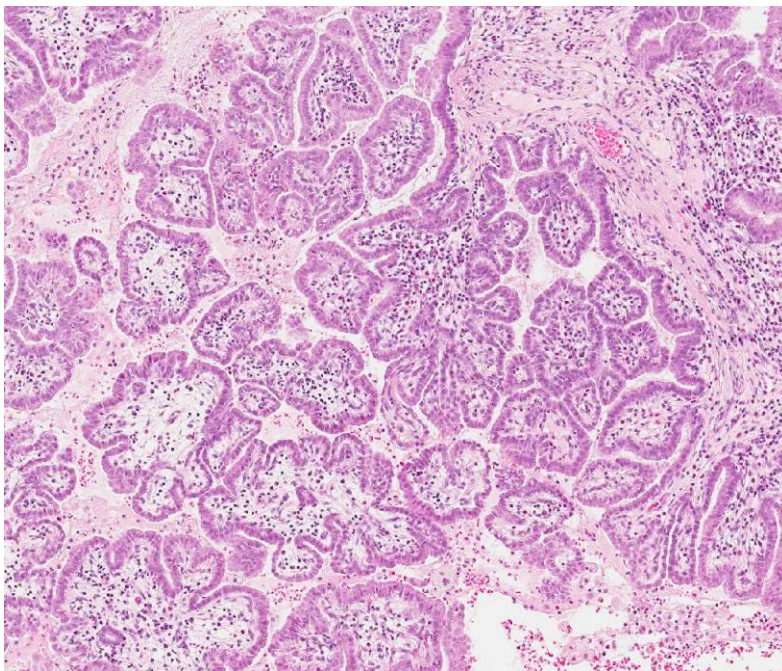
## **1.2 Types, Subtypes and Stages of Lung Cancer**

There are two major types of lung cancer classified as small cell lung cancer (SCLC), and non-small cell lung cancer (NSCLC), which were historically grouped together due to their similar prognosis and management. NSCLC accounts for up to 85% of cases and is generally split into two major subtypes; squamous cell carcinoma and adenocarcinoma, which are complemented by a large number of other rarer sub-types e.g. large cell carcinoma, and pleomorphic carcinomas.

Currently, squamous cell carcinoma is the most common NSCLC type seen in the UK and it is usually observed in smokers. It originates from the squamous epithelial cells that line the airways, often at the lung hila (above and behind the cardiac impression) and in the bronchial bifurcation (the point where the two bronchi split from the trachea) (Figure 1.2). Adenocarcinoma also develops in the cells that line the airways; invariably the mucus secreting cells e.g. goblet cells. Consequently it is often found in the peripheral areas of the lung (Figure 1.3). Adenocarcinoma however is becoming as common as squamous cell carcinoma in the UK due to the increasing prevalence of smoking of “low tar” cigarettes. In other Western countries such as the USA, adenocarcinoma is the most frequent histological subtype, whereas in Eastern countries (e.g. China) squamous carcinoma still predominates.

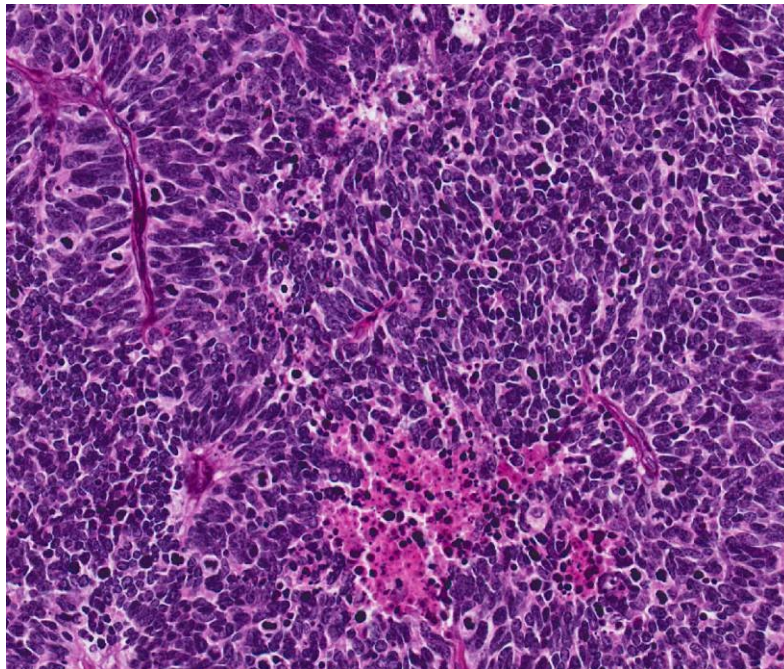


**Figure 1.2** 40x magnification of Squamous Cell Carcinoma of the lung. Image reproduced courtesy of Professor Andrew Nicholson, Consultant Histopathologist, Royal Brompton and Harefield Hospitals NHS Foundation Trust.



**Figure 1.3** 40x magnification of Adenocarcinoma of the lung, showing neoplastic cells arranged in papillary groups or glands. Image reproduced courtesy of Professor Andrew Nicholson, Consultant Histopathologist, Royal Brompton and Harefield Hospitals NHS Foundation Trust.

SCLC accounts for 12% of lung cancers and is predominantly related to smoking. The cells are small and contain dense neurosecretory granules that are visible under the microscope (Figure 1.4). Consequently it has historically been known as “oat cell” cancer due to its appearance. It generally grows rapidly, metastasizes to lymph nodes and distant organs early in disease progression, sooner than is usually observed for NSCLC, and so tends to be advanced at the point of diagnosis. This, together with its initial exquisite chemosensitivity, means that chemotherapy is the most common treatment of SCLC (see Section 1.6). It is frequently associated with distinct endocrine/paraneoplastic syndromes as a result of the neuroendocrine hormones produced by the tumour.



**Figure 1.4** 200x magnification of Small Cell Carcinoma of the lung. Image reproduced courtesy of Professor Andrew Nicholson, Consultant Histopathologist, Royal Brompton and Harefield Hospitals NHS Foundation Trust.

NSCLC is classified into four stage groupings to classify differing prognostic groups, referred to as stages. The currently used version is the Union for International Cancer Control (UICC) version 7 stage grouping (Goldstraw *et al.*, 2007). Stage I is only located in the lungs with no involvement of the lymph nodes. Stage II is in the lung but involves nearby lymph

nodes or large node negative tumours. Stage III is also described as locally advanced disease, and is classified by a heterogeneous group of tumour findings: stage IIIA include tumours which have spread only to lymph nodes on same side of mediastinum as the primary cancer, as well as tumours without mediastinal lymph node involvement when these are either large, involve other lung lobes on the same side, or invade mediastinal structures; and stage IIIB where the cancer may have spread to mediastinal lymph nodes on the opposite side of the primary tumour, or in the supraclavicular fossa (nodes just above the collar bone), or involves the mediastinal nodes on the same side of the tumour as the primary but with a tumour invading the mediastinal structures. Stage IV is the most advanced stage of cancer and is described as advanced disease. The cancer has spread to both lungs or to another part of the body such as the liver or other organs, or causes a pleural or pericardial effusion. The UICC staging also uses the tumour/lymph node/metastases (TNM) system to define the stage of disease. The tumour (T) descriptor can range from T0, where there is no evidence of a primary tumour, to T3, where the tumour may be large (over 7cm in diameter), or T4 indicating invasion into its surrounding structures, or that it has caused different nodules in a different lobe of the same lung. The lymph node descriptor (N) indicative of involvement may range from N0, no regional node metastasis, to N3, where there is involvement of opposite sided mediastinal lymph nodes or nodes in the supraclavicular fossa or scalene nodes. Metastasis can either be M0, no distant metastasis, or M1a, presence of a malignant pleural or pericardial effusion, or lung metastases in the opposite lung, or M1b, presence of distant metastases e.g. in the bones or liver. The TNM descriptors are grouped into stage groupings (Stage I-IV) (Table 1.1).

| Proposed TNM Stage Groupings |        |        |        |
|------------------------------|--------|--------|--------|
| Occult Carcinoma             | TX     | N0     | M0     |
| Stage 0                      | Tis    | N0     | M0     |
| Stage IA                     | T1a, b | N0     | M0     |
| Stage IB                     | T2a    | N0     | M0     |
| Stage IIA                    | T1a, b | N1     | M0     |
|                              | T2a    | N1     | M0     |
|                              | T2b    | N0     | M0     |
| Stage IIB                    | T2b    | N1     | M0     |
|                              | T3     | N0     | M0     |
| Stage IIIA                   | T1, T2 | N2     | M0     |
|                              | T3     | N1, N2 | M0     |
|                              | T4     | N0, N1 | M0     |
| Stage IIIB                   | T4     | N2     | M0     |
|                              | Any T  | N3     | M0     |
| Stage IV                     | Any T  | Any N  | M1a, b |

**Table 1.1** Tumour/Node/Metastasis Staging in lung cancer. As tumour size and local invasion increases (T0 to T4), the stage of disease increases too. If nodal involvement increases (N0 to N3), smaller tumours may lead to a more advanced stage of disease. Once distant metastasis has been observed (M1a or M1b), the cancer is classed as Stage IV. Reproduced with permission from Lippincott Williams and Wilkins/Wolters Kluwer Health: Journal of Thoracic Oncology (Goldstraw *et al.*, 2007), copyright (2007).

SCLC was previously clinically classified as either limited stage or extensive stage by the Veterans Administration Lung Study Group (VASLG) staging system. Here, in limited stage SCLC, the tumour does is not associated with distant metastases, or pleural or pericardial effusion, and is encompassable by a single radical radiotherapy field, usually involving one side of the lung and associated lymph nodes. Extensive stage refers to tumour spread to other regions of chest not encompassable within a radical radiation field or parts of body (Zelen, 1973). This system however has now been updated by the UICC version 7 staging system for SCLC, with the latter being based on a TNM classification akin to that for NSCLC (Valli  res *et al.*, 2009).

### 1.3 Causes of Lung Cancer

In lung cancer, smoking is the primary risk factor and tobacco is estimated to cause almost 90% of all deaths (Parkin, 2011). More than 60 years ago, case-control studies between smokers and non-smokers with the incidence of lung cancer left no doubt that there is a direct association between smoking and carcinoma of the lung (Doll and Hill, 1950, Levin *et al.*, 1950, Wynder and Graham, 1950). The carcinogenic chemicals in tobacco smoke, such as tar, arsenic, hydrogen cyanide, carbon monoxide, benzene and formaldehyde, are known to damage the tissues of the airways and lungs, as well as inducing mutations in the DNA of their cells. Several years after cessation of smoking, lung cancer risk decreases (Cogliano *et al.*, 2011). Other chemical exposures such as asbestos, passive smoking, metal dust, and diesel exhaust fumes have also been linked to lung cancer risk (Cogliano *et al.*, 2011).

After smoking, radon, a naturally occurring gas found in granite (Health Protection Agency, <http://www.hpa.org.uk>), is the second most important risk factor for lung cancer. Other possible risk factors include previous lung disease, such as tuberculosis or pulmonary fibrosis, lowered immunity, and nutritional and dietary patterns (<http://www.cancerresearchuk.org>, Goldstraw *et al.*, 2011, van Meerbeeck *et al.*, 2011). Studies have also shown increased risk if there is prior family history of lung cancer in particular a first-degree relative (Scagliotti *et al.*, 2009b) and this is independent of smoking and other recognized risk factors (Coté *et al.*, 2012). Genetic variation in certain germline genes such as those involved in carcinogen metabolism, DNA damage repair, cell cycle control, and apoptosis, is also thought to have an influence on lung cancer risk (see Sections 1.4 and 1.5 below).

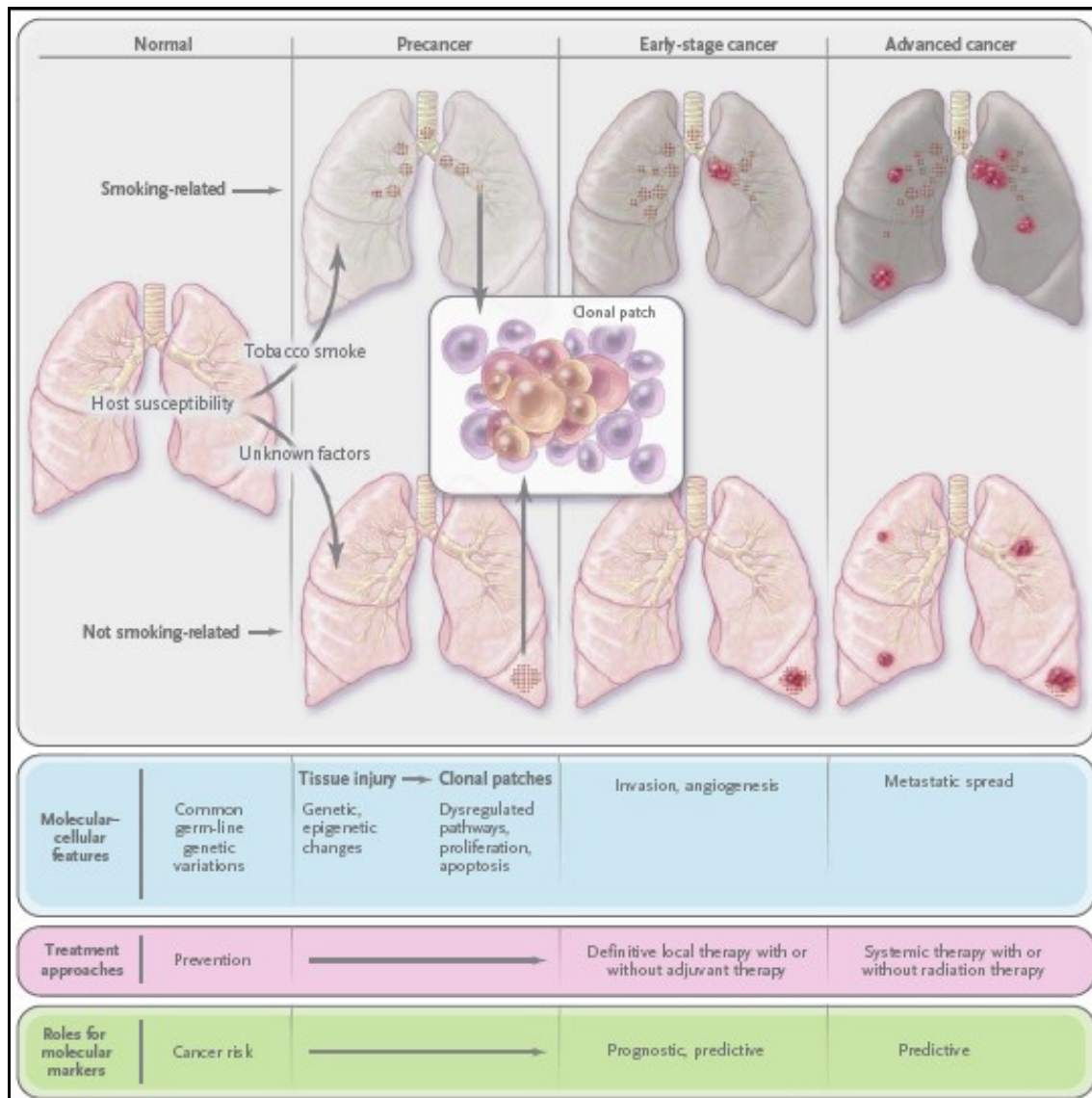
## 1.4 Lung Cancer Biology

Cancer arises due to the mutation of genes (see Section 1.5) leading to various hallmarks of cancer including: self-sufficiency of growth signals, insensitivity to anti-growth signalling, apoptosis evasion, limitless replicative ability, angiogenesis control, invasion and metastasis (Hanahan and Weinberg, 2000). These carcinogenesis-driving mutations generally occur in either oncogenes or tumour suppressor genes (TSGs).

Oncogenes are the altered versions of proto-oncogenes: when a proto-oncogene is mutated or when there is a change in their expression patterns, neoplastic transformation can occur. These mutations and expression changes tend to act in a dominant fashion, where a single alteration in one copy can promote uncontrolled growth. Oncogenes are often growth factors or their receptors (such as epidermal growth factor [EGF]) or their downstream signalling molecules (such as Kirsten Rat Sarcoma Viral Oncogene Homolog [KRAS]).

TSGs are normal cellular genes that generally operate in a recessive manner where loss of both functioning copies is required for tumorigenesis (Knudson's two hit hypothesis). They are often genes that have some control over the cell cycle or those involved in DNA repair mechanisms, such as nucleotide excision repair, mismatch repair and base excision repair. *p53* is a key example of a TSG that is mutated in a multitude of cancers, from colon and breast cancers through to leukemia and lymphomas, and inherited mutations in the gene can lead to Li-Fraumeni Syndrome. This leads to increased susceptibility to a variety of cancers, including lung cancer, and is passed from parent to child in an autosomal dominant fashion. Lung cancer evolves due to the interaction of both environmental and genetic susceptibility factors that influence carcinogenesis. Unrelated factors such as hormonal and viral factors may also play an important role (see Section 1.3).

Smoking-related cancers often develop in the central airway, the primary point of tissue injury, whereas in most tumours unrelated to smoking tissue injury occurs and develops in the peripheral airway. This injury leads to genetic and epigenetic changes, such as base changes, insertions, deletions and promoter methylation. These changes can result in alterations in the transcriptome, such as the activation of inflammation and apoptosis pathways, or more local changes in particular gene expression of growth factor pathways e.g. the epidermal growth factor receptor (EGFR) pathway. In addition, mutation rates also increase in the presence of substantial exogenous mutagenic exposures such as tobacco smoke, naturally occurring chemicals such as aflatoxins (related to liver cancer) or radiation (UV) (Stratton *et al.*, 2009). These changes may persist for some time and eventually lead to aberrant pathway activation and cellular function to produce premalignant changes, including dysplasia and clonal patches (Figure 1.5). As this continues, further alterations can lead to invasion and angiogenesis and thus early-stage cancer eventually leading to advanced cancer and metastasis (Herbst *et al.*, 2008).



**Figure 1.5** Diagram illustrating the evolution of lung cancer that is smoking-related (top) and non-smoking-related (bottom). The diagram indicates the potential localization of the cancer mass by stage of the disease as well as providing details of the molecular and cellular features, treatment approaches and the potential role of molecular markers. Reproduced with permission from (Herbst *et al.*, 2008), Copyright Massachusetts Medical Society.

Lung cancers that are related and unrelated to smoking (such as SCLC and NSCLC adenocarcinoma respectively) have very different somatic molecular profiles as exemplified by the difference in the mutational status of a number of key genes i.e. *p53*, *KRAS*, *EGFR* and erythroblastic leukemia viral oncogene homolog 2 (*ERBB2*, also known as *HER2*).

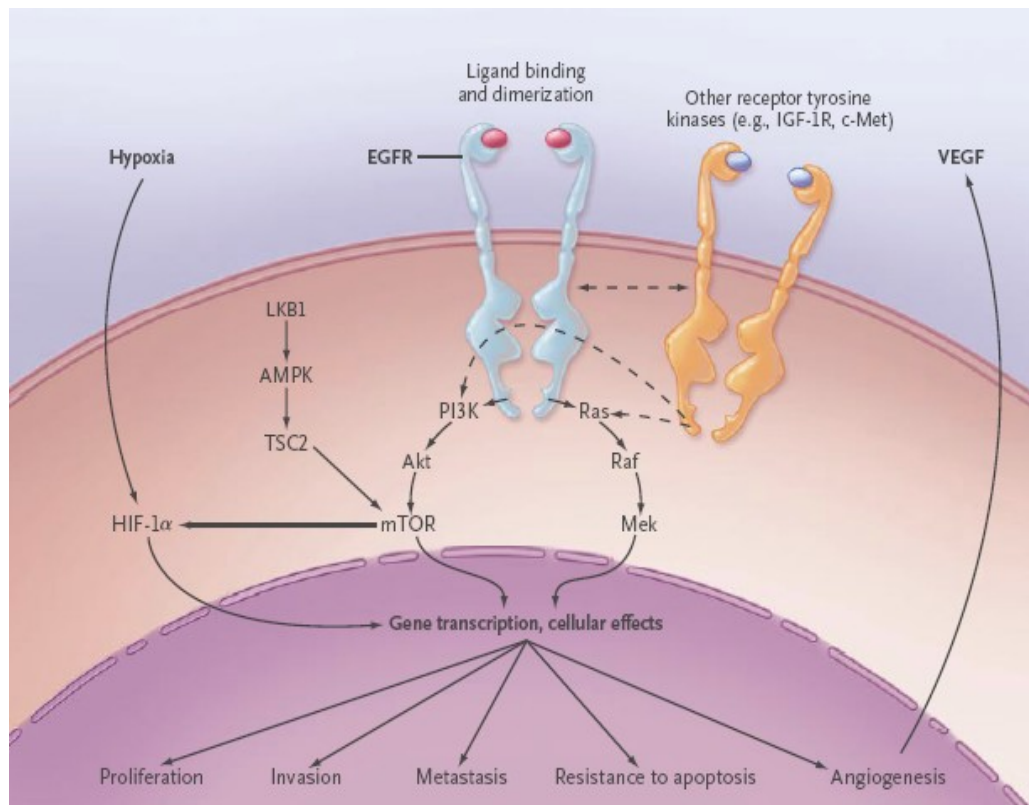
A biomarker is a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathological processes or pharmacological responses to a therapeutic intervention. Within the tumour, these characteristics, such as a DNA sequence, gene expression signature or copy number variant, can be objectively measured or evaluated as an indicator of a given phenotype, for example, a pharmacological response in a disease versus normal state. Biomarkers may be predictive of risk, prognosis, and sensitivity to treatment and can work in addition to the traditional staging system (see Section 1.2). They may be stage specific, spanning the whole disease course from early stages through to the later stages and may define mechanisms of therapy resistance (Herbst *et al.*, 2008). As they are measurable characteristics they are clinically useful. A simple test, such as genotyping a tumour sample, may quickly indicate sensitivity or resistance to a particular treatment (a predictive biomarker). With biomarker directed therapy, it is possible to personalize treatment, giving the correct treatment to the correct patient at the correct time. The biomarker may not predict any therapeutic utility but indicate prognosis (prognostic biomarker e.g. KRAS mutations are associated with poorer prognosis). Some biomarkers can be both predictive and prognostic (e.g. EGFR mutations predict superior outcomes with EGFR-directed therapies and also independently confer an improved prognosis compared with EGFR wild-type tumours).

## 1.5 Genetics, Genomics, Transcriptomics and Pharmacogenomics

Mutations occur sporadically in carcinogenesis and these often give the host cell key characteristics that allow disease progression, sometimes referred to as the hallmarks of cancer. These characteristics include self-sufficiency in growth signals, irresponsiveness to anti-growth signals, evasion of apoptosis, limitless replicative potential, angiogenesis, and invasion and metastasis (Hanahan and Weinberg, 2000). Mutations may also affect response to chemotherapy. Possible mutations include simple point mutations, amplification and/or overexpression of genes, changes in gene copy number and structural rearrangements – Mutant Allele Specific Imbalances (MASI). Oncogenes, such as *KRAS*, *EGFR* and other genes in similar pathways (Figure 1.6), are recognized to be of particular importance, especially in response to chemotherapeutics. Many genes have been shown to play a role in NSCLC carcinogenesis, both as substantial players or drivers, as part of coordinated control, or as potential susceptibility loci.



or heterodimer formation with closely related molecules such as ERBB2 (Human Epidermal Growth Factor Receptor 2, also known as HER2). Auto-phosphorylation and trans-phosphorylation of the receptors through their tyrosine kinase domains (exons 18-24 in *EGFR*) results in downstream effector recruitment and activation of proliferative and cell-survival signals (Lynch *et al.*, 2004) (Figure 1.7).



**Figure 1.7** Schematic showing dimerisation of EGFR and its signalling pathways, PI3K/AKT and RAS/RAF, and their influence on downstream cellular functions. Reproduced with permission from (Herbst *et al.*, 2008), Copyright Massachusetts Medical Society.

*EGFR* over-expression has been observed in the development of NSCLC in 40-80% of cases (Lynch *et al.*, 2004, Herbst *et al.*, 2008) and somatic *EGFR* mutations are thought to be early events in tumorigenesis. These are more frequently found in female adenocarcinoma patients and patients of East Asian origin (30-50% of East Asians compared to 10% of Americans). *EGFR* rearrangements in the extracellular domain of the molecular are rare (Lynch *et al.*, 2004). Somatic mutations are usually found within the *EGFR* tyrosine kinase domain, often

near the ATP cleft, where ATP binds to enable phosphorylation (Lynch *et al.*, 2004), and most mutations in adenocarcinomas involve in-frame deletions in exon 19 or substitutions in exon 21 (L858R) (Lynch *et al.*, 2004, Herbst *et al.*, 2008). The heterozygous nature of *EGFR* mutations is evidence for its dominant oncogenic effects as the mutations are not found in normal lung tissue DNA (Paez *et al.*, 2004) providing evidence for their somatic acquisition. These mutations are also seen more commonly in never-smoking related cancers (Sholl *et al.*, 2009) and in comparatively younger age groups (less than 55 years old) (Gandhi *et al.*, 2009).

*EGFR* mutations are known to be markers for therapy choices as they may lead to truncated proteins that are sensitive to tyrosine kinase inhibitors (TKIs), such as erlotinib and gefitinib. These biological small molecules act by inhibiting the ATP cleft within the tyrosine kinase domain (Lynch *et al.*, 2004) by competing out ATP and thus block the proliferative action of the downstream activation of PI3K (phosphoinositide 3-kinase)/AKT signalling. The predictive value of *EGFR* mutations for use of gefitinib has been strengthened by the results of seven randomised Phase III trials that specifically compared TKIs used as a first-line therapy with traditional platinum-based therapy (see Section 1.6) in patients with advanced NSCLC. Both progression free survival and response rates for gefitinib were improved compared to platinum-based chemotherapy in the first-line setting of metastatic *EGFR*-mutated NSCLC (Antonicelli *et al.*, 2013).

Resistance to TKIs however always occurs. This resistance can be intrinsic or more usually develops after TKI treatment (acquired resistance) (Engelman *et al.*, 2007). Acquired resistance has been demonstrated to occur through two major mechanisms. A second mutation (T790M) encoded by exon 20 of the *EGFR* gene and within the tyrosine kinase domain of the molecule has been shown to confer resistance to *EGFR* TKIs, in the presence of either wild-type (wt) or mutant *EGFR*. Tumours harbouring this double mutant appears to

have the same resistance to tyrosine kinase inhibitors as those with wt EGFR (Pao *et al.*, 2005a). Germline transmission of this mutation has also been reported (Bell *et al.*, 2005). The other commonest mechanism is the recruitment of other kinases. For example amplification of another receptor tyrosine kinase proto-oncogene MET (mesenchymal epithelial transcription factor) has also been shown to render susceptible EGFR mutants resistant to TKIs (Engelman *et al.*, 2007). This happens by initiating PI3K/AKT signalling and its downstream proliferative effects via ERBB3. Intrinsic resistance in EGFR mutants may happen through insertions in exon 20 of *EGFR*, loss of the *PTEN* gene, activation of MET, absence of EGFR expression or mutations in *HER2* tyrosine kinase (Toschi and Cappuzzo, 2010). *PTEN* homozygous deletion and the absence of its protein confers resistance to TKIs, again by PI3K/AKT pathway interference (Gandhi *et al.*, 2009). TKI sensitivity has also been significantly associated with *EGFR* copy number gains *in vitro* although *in vivo* evidence is conflicting (Gandhi *et al.*, 2009, Toschi and Cappuzzo, 2010) (see Section 1.5.8).

MET also forms heterodimers with EGFR, HER2, HER3 and RET. These other ERBB family proteins may additionally have a key role in cancer themselves their depletion may induce apoptosis in association with AKT/ERK signalling inhibition as well as reduction of cell migration and STAT3 signalling (Tanizaki *et al.*, 2011). Other methods of acquiring TKI resistance include increased signalling through parallel pathways such as VEGF and IGF1R, mutation and activations of PIK3CA, and transformation into a small-cell lung cancer phenotype (Antonicelli *et al.*, 2013).

The prevalence of *EGFR* mutations in NSCLC tumours ranges from approximately 15% in Caucasians to almost 50% in East Asians, with the majority (95%) of these being found in adenocarcinomas (Antonicelli *et al.*, 2013).

### 1.5.2 KRAS

The *KRAS* gene is located on the reverse strand of the short arm of chromosome 12, is nearly 6,000 base pairs long, has four coding exons and the resultant protein consists of 188 amino acids. Somatic *KRAS* mutations are prevalent in many cancers, ranging from colorectal and pancreatic cancer to leukemia (Loriot *et al.*, 2009). In NSCLC, activating mutations are generally limited to adenocarcinoma and are almost always mutually exclusive of *EGFR* mutations (Pao *et al.*, 2005b, Herbst *et al.*, 2008, Adams and Harvey, 2010) and *BRAF* mutations (Soh *et al.*, 2009). *KRAS* activating mutations however are one of the most common molecular changes in NSCLC (Roberts and Stinchcombe, 2013). The mutations are more commonly seen in smokers in contrast to *EGFR* where mutations are seen in non-smokers (Sholl *et al.*, 2009). Mutations are often smoking related G>T transversions (Herbst *et al.*, 2008) and hotspots have been found in the GTPase domain of the gene in codons 12, 13 and 61 (Loriot *et al.*, 2009). This area is key for the function of the protein; the GTP hydrolysis performed by this domain leads to the negative regulation of signal transduction. Mutations in this area can lead to interference with the GTPase activity and reduced negative regulation. Downstream effectors of *KRAS*, such as *BRAF*, *MEK1 & 2* and *ERK1 & 2*, may also be mutated. *MEK* and *ERK* gene amplifications however have not been observed in lung cancer.

*KRAS* mutations tend to be classically associated with the adenocarcinoma histology (mutated in 20-30% [Roberts and Stinchcombe, 2013]). In contrast to *EGFR* mutations, *KRAS* mutations are more frequent in male, smoking patients and are also less commonly identified in other NSCLC subtypes such as squamous carcinomas and large cell carcinoma (Gandhi *et al.*, 2009).

*KRAS* mutated tumours have intrinsic resistance to tumour response to EGFR TKIs such as gefitinib or erlotinib (Pao *et al.*, 2005b, Gandhi *et al.*, 2009), although there is some evidence for prolonged disease stabilization under this therapy. This is not surprising since *KRAS* sits downstream of EGFR and therefore inhibiting upstream EGFR would not be predicted to benefit. It is possible however, that these particular *KRAS* mutants may have increased *EGFR* copy number (Toschi and Cappuzzo, 2010). *KRAS* mutational status in general appears to be associated with poor survival, prognosis (Mascaux *et al.*, 2005) and diminished response to most systemic therapies (Adams and Harvey, 2010). Despite this it may still be a useable tool in guiding therapy, for example whether a targeted agent or conventional chemotherapy should be used (Loriot *et al.*, 2009).

*BRAF* (v-raf murine sarcoma viral oncogene homolog B1), a downstream effector of *KRAS*, is a serine threonine kinase enzyme that links RAS GTPases with the MAPK family, the controllers of cell proliferation. Mutations in *BRAF* have been well documented in melanomas, with the majority (80%) of mutations affecting exon 15 with a missense mutation Val600Glu (V600E). Lung cancers harbouring *BRAF* mutations (approximately 1-3% of adenocarcinomas) have a more diverse range of mutations, although V600E is by far the most common variant observed. There are very few instances of the missense mutation Val600Glu but often Leu596Val and Gly468Ala in the kinase domain and activation domain respectively are found. *BRAF* mutations also appear to be generally mutually exclusive of *KRAS* and *EGFR* mutations (Vijayalakshmi and Krishnamurthy, 2011).

### 1.5.3 ALK and Other Gene Rearrangements

The somatically acquired translocation of involving the genes echinoderm microtubule-associated protein-like 4 (*EML4*) and anaplastic lymphoma kinase (*ALK*) is highly specific to NSCLC and was first described in 2007 (Soda *et al.*, 2007, Shaw *et al.*, 2009). It is caused by an inversion on chromosome 2p leading to a fusion gene of the N-terminal of *EML4* and the C-terminal of *ALK* leading to expression of a chimeric tyrosine kinase. *EML4* is found on the forward strand of the short arm of chromosome 2, consists of 5,549 base pairs with 23 exons and with a protein of 981 amino acids. *ALK* is also found on the short arm of chromosome 2 but on the reverse strand and consists of 6,220 base pairs, 29 coding exons with a protein of 1,620 amino acids.

It is suggested that the basic domain within *EML4* may mediate dimerisation of *EML4-ALK*. There is also some indication for diversity in the *EML4* breakpoint region, which may or may not have an effect on this dimerisation capability and hence tyrosine kinase activity (Soda *et al.*, 2007).

The translocation has been detected in both smokers and non-smokers and is distinct from *EGFR* mutations. It tends to be observed in young never smokers, more frequently male, and *ALK* positive tumours usually have signet ring morphology by light microscopy. This morphology however can be difficult to see with the small biopsy material that is usually used for diagnostic purposes. The cells of the tumour contain large quantities of mucin, which pushes the nucleus to the edge of the cell, giving them the appearance of a signet ring. A number of other *ALK* fusion partners have also been identified (Shaw and Engelman, 2013). The translocation tends to be mutually exclusive of both *EGFR* and *KRAS* mutations. Patients are generally younger and do not benefit from TKI treatment, possibly due to constitutive *ALK* signalling (Toschi and Cappuzzo, 2010). Response to chemotherapy is generally similar

to tumours harbouring wt ALK and EGFR, although pemetrexed does demonstrate increased activity. The constitutional activation of the ALK kinase conferred by the oncogenic fusion can however be effectively blocked using small molecule inhibitors of ALK e.g. crizotinib with dramatic clinical benefits (Shaw *et al.*, 2009).

In addition to the rare somatic fusion protein, EML4-ALK, others have been seen as well, granted less frequently. The translocation of kinesin family member 5B (KIF5B) to the ret proto-oncogene (RET) has been seen exclusively in adenocarcinomas that harbour neither an *EGFR* nor a *KRAS* mutation (Takeuchi *et al.*, 2012). A number of RET inhibitors are licensed or in development, and trials of RET kinase inhibitors e.g. vandetanib in NSCLC are ongoing. Another rare somatic fusion is the translocation of the c-ros oncogene receptor tyrosine kinase (ROS1). Although little is known about the roles of wild-type ROS1, oncogenic gene fusions involving its 3-prime region have been identified in various human cancers including NSCLC (Davies *et al.*, 2012). These ROS1 fusions have principally observed in adenocarcinomas (Takeuchi *et al.*, 2012). Due to sequence homology between ROS and ALK kinase, ALK kinase inhibitors show efficacy against ROS1 fusion adenocarcinomas and an ongoing trial of crizotinib in ROS1 fusion NSCLCs has demonstrated remarkable efficacy.

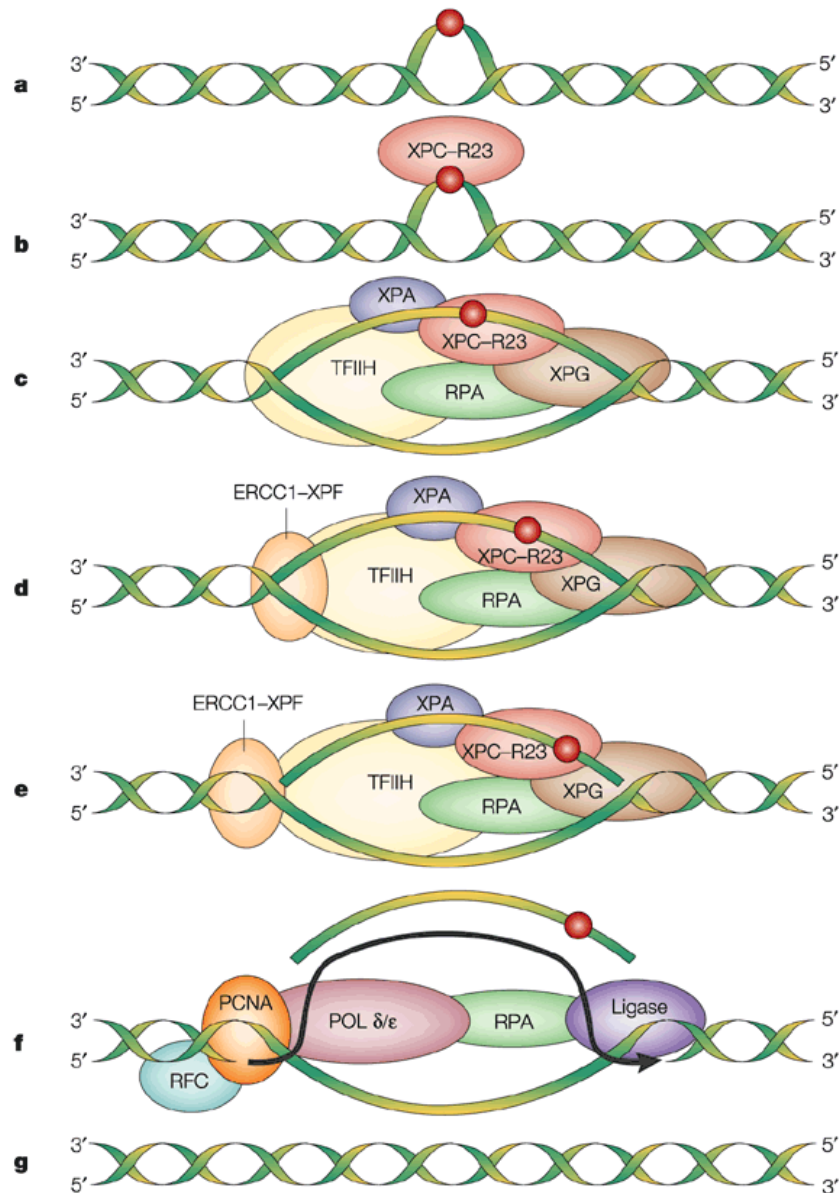
#### 1.5.4 Developmental Genes: *PAX9*, *TTF1*, *NKX2-8* and *Shh*

Genes involved in lung development by nature of their function have undergone prior investigation and been implicated in tumorigenesis. Genes have included those encoding the transcription factors thyroid transcription factor (TTF-1), NK2 homeobox 8 (NKX2-8) and Paired Box 9 (PAX9) (Hsu *et al.*, 2009), all three being located on chromosome 14q13 (Herbst *et al.*, 2008, Hsu *et al.*, 2009). The co-activation of TTF-1 and NKX2-8 has been associated with clusters of patients with poor survival. Cell lines harbouring this co-activation have been shown to exhibit resistance to cisplatin doublet chemotherapy as well as to taxanes, gemcitabine and vinorelbine, reflecting the importance of understanding genetic signatures and cooperation between multiple genes and their expression patterns in cancer. These signatures representing transcription factor statuses have both prognostic and predictive value and may give indications for individual treatment and therapeutic options (Hsu *et al.*, 2009).

*Sonic Hedgehog (Shh)* is another gene involved in lung development and consequently may have a role in regeneration and carcinogenesis. After acute injury of the airways, the hedgehog pathway is extensively activated in the epithelium. It has been suggested that epithelium progenitors that retain Hh signalling may progress into a SCLC malignancy (Watkins *et al.*, 2003). This may indicate a possible treatment by blockade of the Hh pathway particularly in cases of SLCLC with poor prognosis (Watkins *et al.*, 2003).

#### 1.5.5 DNA Repair and Nucleotide Excision Repair (NER): *ERCC1*, *RRM1* and *BRCA1*

When an exogenous agent causes damage to DNA, the structure and chemistry of the DNA duplex undergoes changes. The nucleotide excision repair (NER) pathway removes this damage by recognizing the alteration through the xeroderma pigmentosum C (XPC) protein, which in turn recruits further proteins (XPA and replication protein A [RPA]) that facilitate specific base damage recognition. The area around the DNA damage is unzipped by helicase activities of XPB and XPD. The heterodimeric sub-complex of Excision Repair Cross-Complementation 1 (ERCC1) – XPF then binds the NER complex and, with XPG, enables the removal of a 27-30 nucleotide fragment of DNA (containing the damaged base) to be removed. The DNA is then repaired by DNA polymerase  $\delta/\epsilon$  with other accessory replication proteins and DNA ligase restores the strand's integrity (Figure 1.8). Through this mechanism of detection, protein recruitment, removal of the damage and re-synthesis, genetic integrity at the base pair level is maintained.



**Figure 1.8 a** | Nucleotide excision repair (NER) operates on base damage caused by exogenous agents (such as mutagenic and carcinogenic chemicals and photoproducts generated by sunlight exposure) that cause alterations in the chemistry and structure of the DNA duplex. **b** | Such damage is recognized by a protein called XPC, which is stably bound to another protein called HHRAD23B (R23). **c** | The binding of the XPC–HHRAD23 heterodimeric sub-complex is followed by the binding of several other proteins (XPA, RPA, TFIIH and XPG). Of these, XPA and RPA are believed to facilitate specific recognition of base damage. TFIIH is a sub-complex of the RNA polymerase II transcription initiation machinery which also operates during NER. It consists of six subunits and contains two DNA helicase activities (XPB and XPD) that unwind the DNA duplex in the immediate vicinity of the base damage. This local denaturation generates a bubble in the DNA, the ends of which comprise junctions between duplex and single-stranded DNA. **d** | The subsequent binding of the

ERCC1–XPF heterodimeric sub-complex generates a completely assembled NER multi-protein complex. **e** | XPG is a duplex/single-stranded DNA endonuclease that cuts the damaged strand at such junctions 3' to the site of base damage. Conversely, the ERCC1–XPF heterodimeric protein is a duplex/single-stranded DNA endonuclease that cuts the damaged strand at such junctions 5' to the site of base damage. This bimodal incision generates an oligonucleotide fragment ~27–30 nucleotides in length which includes the damaged base. **f** | This fragment is excised from the genome, concomitant with restoring the potential 27–30 nucleotide gap by repair synthesis. Repair synthesis requires DNA polymerases  $\delta$  or  $\epsilon$ , as well as the accessory replication proteins PCNA, RPA and RFC. The covalent integrity of the damaged strand is then restored by DNA ligase. **g** | Collectively, these biochemical events return the damaged DNA to its native chemistry and configuration. ERCC1, excision repair cross-complementing 1; PCNA, proliferating cell nuclear antigen; POL, polymerase; RFC, replication factor C; RPA, replication protein A; TFIIH, transcription factor IIH; XP, xeroderma pigmentosum. Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Cancer (Friedberg, 2001), Copyright 2001.

Aberrations in DNA repair pathways are well documented in several cancers, from colon cancers with microsatellite instability to breast cancer with inherited defective *BRCA* genes, the latter being members of the Homologous Recombination pathway. Alterations in NER have also been documented in lung cancer.

*Excision Repair Cross-Complementation (ERCC1)* is located on chromosome 19q and is involved in the process of NER (Gazdar, 2007). Increased ERCC1 protein expression has been associated with lack of response to platinum-based chemotherapy (Adams and Harvey, 2010) and poor survival in cisplatin-treated patients (Rosell *et al.*, 2003). In contrast patients with low expression levels or an absence of ERCC1 tend to have a prolonged median survival time and improved response to platinum chemotherapy. There is some evidence suggesting that the effect may be sex specific where men with ERCC-negative tumours treated with carboplatin and gemcitabine had a significantly increased survival compared with those harbouring ERCC-positive tumours (Chen *et al.*, 2010). Down-regulation of the complex of

ERCC1 and XPF (another NER protein) has also been shown to result in sensitization to cisplatin (Arora *et al.*, 2010). Recent data however has indicated that platinum resistance due to ERCC1 expression is conferred by a specific ERCC1 isoform that is difficult to measure using standard methodology (Friboulet *et al.*, 2013). As such most studies measuring ERCC1 are likely to be strongly biased.

Ribonucleotide reductase M1 (RRM1) is an enzyme that is essential for deoxyribonucleotide production prior to DNA synthesis in the cell cycle and NER. It has been shown to predict response to therapeutic agents, with high levels of protein expression leading to chemotherapy (cisplatin and gemcitabine) resistance (Rosell *et al.*, 2003, Adams and Harvey, 2010,). In contrast to ERCC1 expression, high co-expression of ERCC1 and RRM1 has been shown to be associated with statistically significant increased disease-free and overall survival in early stage (Stage I) NSCLC patients. *ERCC1* and *RRM1* are not molecular predictors of efficacy to the same extent as *EGFR* mutations or *ALK* rearrangements (Besse *et al.*, 2013).

BRCA1 is involved in transcription coupled repair (Friedberg, 2001) and is well known to play a role in both hereditary and sporadic breast and ovarian cancers. Low BRCA1 expression in cisplatin-treated NSCLC patients has been associated with prolonged, progression-free survival whilst poorer median survival associates with high BRCA1 levels (Toschi and Cappuzzo, 2010). This implied involvement of BRCA1 in NSCLC chemoresistance is however based on observations from studies of ovarian and breast cancers and not NSCLC *per se* (Toschi and Cappuzzo, 2010). Activation of poly(ADP-ribose) polymerase-1 (PARP-1) in response to DNA damage is an important mechanism to keep homeostasis or to trigger apoptosis (Ahmad *et al.*, 2010). PARP inhibitors have already shown potent anti-tumour activity (Ashworth, 2008) in both hereditary and sporadic breast

and ovarian cancers that are deficient in BRCA1, which has led to the suggestion that those NSCLC tumours with sporadic BRCA1 loss could also benefit from PARP inhibition (Paul *et al.*, 2010) and trials of these inhibitors in NSCLC and SCLC are ongoing.

Another DNA repair pathway is non-homologous end joining (NHEJ). This mends double-stranded breaks (DSB) in DNA and helps prevent large chromosomal changes within the cancer genome. It has been shown that cisplatin-induced DNA lesions that are physically close to the termini of DSB (within 6 base pairs) can inhibit rejoining (Sears and Turchi, 2012) which it could be implied may lead to further gross chromosomal changes.

#### 1.5.6 Cell Cycle Regulators

Genes as well as proteins involved in the regulation of the cell cycle have also been associated with NSCLC. In one study (Weir *et al.*, 2007) two significant focal deletions in well documented *TSGs*: *CDKN2A* and *CDKN2B* were observed in patients. *CDKN2A* and *CDKN2B* are both involved in the inhibition of two cyclin-dependent kinases, CDK4 and CDK6, whose amplification has also been observed (Weir *et al.*, 2007). Another homozygous deletion in *CDKN2A* was observed in a separate study (Ikediobi *et al.*, 2006) that also interestingly found a loss in methylation profile of *CDKN2A*.

The wide reach of cell cycle regulators may mean that they are also involved in or feed into other important molecular pathways, some of which have been discussed already (Sections 1.5.1 and 1.5.2), and may affect lung cancer indirectly, such as *EGFR*, *KRAS* and *BRAF* (Eymin and Gazzeri, 2010). It is possible that they may be of use in combined therapies or prognoses.

### 1.5.7 Other Miscellaneous Somatic Mutations

Certain genes may harbour rare somatic mutations in NSCLC tumours. These include various mutations and Copy Number Gains (CNGs) in the fundamental, cell development and growth *PI3K* signalling pathway. These have tended to be observed for squamous cell carcinomas (Yamamoto *et al.*, 2008, Gandhi *et al.*, 2009, Adams and Harvey, 2010). One of the PI3K subunits (*PIK3CA*) is on a chromosomal location (3q26) that is frequently amplified in lung cancers (Vijayalakshmi and Krishnamurthy, 2011). Adenocarcinomas with *PIK3CA* gains (approximately 20% [Wojtalla and Arcaro, 2011]) tend to also have other gene mutations suggesting further mutations may be required for progression.

Increased activation of another molecule (by amplification, over-expression, or over-activation) involved in the PI3K pathway, v-akt murine thymoma viral oncogene homolog 1 (*AKT*) has also been seen. This, based on immunohistochemical investigations of tumour samples, appears to be more balanced between the squamous cell and adenocarcinoma subtypes (68% versus 62% respectively (Wojtalla and Arcaro, 2011)) than *PIK3CA* mutations. Over-expression of *AKT* has also been associated with low expression of the TSG: phosphatase and tensin homolog (*PTEN*). This is an important negative regulator of the PI3K/AKT signalling pathway and its loss through homozygous deletion or mutation is well documented in many different cancers although genetic alterations are rare in NSCLC. Loss or reduced expression of the PTEN protein, potentially through transcriptional or translational control, has been reported (Kokubo *et al.*, 2005). The molecular mechanism underpinning this observation however remains unclear but, if elucidated, it could offer further therapeutic targets. Alterations in this pathway are an obvious target for therapy especially at the start of tumour progression as they appear to happen early on. There are several potential gene targets within the pathway (*PI3KCA*, *AKT*, and the mammalian target of rapamycin [*mTOR*])

that could be used to alleviate the pro-survival signals in the cancer and induce apoptosis (Figure 1.6 and Figure 1.7).

Serine/threonine kinase 11 (STK11, also known as LKB1) may have a role in premalignant lesion development as it also interacts with the PI3K/AKT pathway and is a negative regulator of mTOR signalling (Gill *et al.*, 2011). Loss of heterozygosity (LOH) or homozygous inactivation of *STK11* has been seen in tumours that subsequently demonstrate more frequent metastasis (Ji *et al.*, 2007). It may work in cooperation with KRAS activation to initiate pulmonary neoplasia and its methylation status may be of importance (Ikediobi *et al.*, 2006, Ji *et al.*, 2007, Herbst *et al.*, 2008).

Mutations in *ERBB2*, another member of the EGFR kinase family, have been reported mostly in a small fraction of adenocarcinomas. Opinion is divided as to who is affected; current and ex-smokers (Stephens *et al.*, 2004, Ikediobi *et al.*, 2006) or similar populations to those with *EGFR* mutations i.e. female non-smokers (Tomizawa *et al.*, 2011). Little difference however has been seen in overall patient survival of *ERBB2* mutants compared to *EGFR* mutants and those patients harbouring wt *EGFR* and wt *ERBB2*. Nonetheless patients with the mutations may benefit from anti-*ERBB2* therapy such as trastuzumab, afatinib or dacomitinib (Tomizawa *et al.*, 2011).

Thymidylate synthetase (TYMS) is an enzyme that is critical in maintaining the dTMP (thymidine-5-prime monophosphate) pool for DNA synthesis and repair. It is considered among the targets of newer anti-folate drugs, such as pemetrexed, while allelic variants of TYMS have been shown to interfere with platinum activity *in vitro*. In many studies, expression is reported as an unfavourable prognostic factor, although occasionally it has been associated with a favourable outcome (Kotoula *et al.*, 2012). Further investigation revealed that aberrant intra-tumoral TYMS predicted shorter progression free survival than normal

(Kotoula *et al.*, 2012). Data from a number of studies are however conflicting and so the clinical utility of this target remains unknown.

The discoidin domain receptor tyrosine kinase 2 (DDR2) is another potential target for lung cancer. Some novel somatic mutations have been reported in a proportion (3.8%) of squamous cell lung cancers and these oncogenic gain-of-function mutations could be potentially important therapeutically. Sensitivity of the mutated tyrosine kinase to dasatinib, a TKI, has been reported although the mechanism of transformation promoted by DDR2 is still unclear (Hammerman *et al.*, 2011). A trial of dasatinib in DDR2 mutant NSCLC is ongoing [reference number NCT01514864].

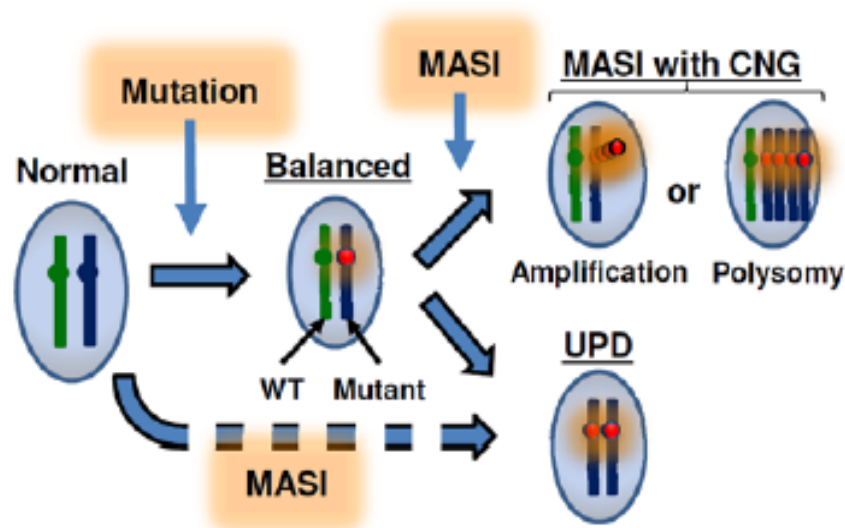
As already mentioned earlier (Section 1.5.1), receptor tyrosine kinases have been shown to play a vital role in lung cancer development. The fibroblast growth factor receptor 1 (FGFR1) is another receptor that has been shown to have increased expression and elevated copy number in some squamous cell carcinomas (approximately 10%) (Kohler *et al.*, 2012). No significant associations, however, between expression and clinico-pathological parameters have been elucidated. Its amplification has already been associated with cancer cell survival and chemotherapy resistance (Kohler *et al.*, 2012). Trials of FGFR kinase inhibitors, e.g. AZD4547, in FGFR-amplified tumours are ongoing [reference number NCT01795768].

Whole-genome linkage analysis also identified CD44, a cell-surface glycoprotein involved in cell-cell interactions, cell adhesion and migration, as important in conferring resistance against carboplatin, another platinum compound that is used in NSCLC chemotherapy (Shukla *et al.*, 2009).

1.5.8 Other Alterations – Structural Rearrangements, Copy Number Alterations, Mutant Allele Specific Imbalances, Uniparental Disomy and Loss of Heterozygosity

As well as somatic mutations (point, insertion or deletion) accumulating during cancer growth, other types of mutation that occur in cancer can affect subsequent gene expression. These include somatic Copy Number Alterations (CNAs) (gain and loss), structural rearrangements, loss of heterozygosity (LOH), uniparental disomy (UPD) and mutant allele specific imbalances (MASI).

MASI may occur after somatic mutation by UPD, amplification of the mutated gene or by polysomy with the mutated gene (Figure 1.9). MASI has been observed in cell lines and tumours harbouring oncogenic mutations. *KRAS* mutations have frequently demonstrated complete or relative MASI (Mitsudomi *et al.*, 1991) and activating *EGFR* exon 19 deletion mutations have been seen in great excess in lung adenocarcinoma (Lynch *et al.*, 2004, Paez *et al.*, 2004). More recently the frequent presence of copy number gains (CNGs) of mutant forms of the same gene has been found in tumour cells.



**Figure 1.9** Three different ways of acquiring MASI – 1) after somatic mutation by UPD where one copy of the chromosome carrying the mutated gene is replicated; 2) where the mutated gene is amplified; or 3) by polysomy of the chromosome carrying the mutated gene. Figure taken from Soh *et al.*, 2009 (PLoS One, 4, e7464: open access journal).

Genome wide approaches such as single nucleotide polymorphism (SNP) arrays have been invaluable in enabling evaluation of dynamic chromosomal and focal changes of CNAs and LOH. Novel lesions have been identified with amplification and/or LOH across several organs. UPD is frequently observed in several cancers. It typically arises when two copies of the same chromosome are inherited from one parent. The acquisition of UPD leads to homozygosity for any pre-existing gene mutations, with selective retention of the mutated allele. Acquired UPD, caused by the loss of one chromosome during cell division followed by the duplication of the remaining homologous pair, is associated with oncogenic sporadic mutations. It has been reported in hematopoietic malignancies (*FLT3* and *WT1* mutations in acute myeloid leukemia; *JAK2* mutations in myeloproliferative disorders [Fitzgibbon *et al.*, 2005]).

To date, acquired UPD in solid tumours has only been observed in TSGs that require two hits to be inactivated (Soh *et al.*, 2009). Most inactivating mutations of TSGs are frequently accompanied by wt allele loss. In one study of cell lines derived from multiple tumour types 20% of activating oncogene mutations were found to be homozygous. MASI detection requires oncogenic mutation identification, analysis of copy number of the mutant gene in tumour cells and determination of the relative ratio of mutant to wt allele (Soh *et al.*, 2009).

CNAs and MASI could play an important role in understanding lung cancer pathogenesis and once more is known about their biological and clinical significance they may provide prognostic markers that indicate tumour and patient response to therapy.

In tumours, CNGs in *EGFR*, *KRAS* and *ERBB2* (Gandhi *et al.*, 2009) have been frequently found (>10%) with gains in *EGFR* nearly twice as common as any other gene. Unlike other mutations, CNGs on the whole appear to have no bias towards any histological subtype with the exception of *PI3K* and *BRAF* gains which appear to be limited to squamous cell

carcinoma and adenocarcinoma respectively. Mutual exclusivity of CNGs has not been found but there is a significant association between mutation and CNG of individual genes. They are not completely random. For example mutant *EGFR* leads to increased likelihood of EGFR gain and mutant *KRAS* leads to increased likelihood of KRAS gain but the mutation must precede the CNG (Gandhi *et al.*, 2009). In addition, MASI is exhibited with the mutant allele almost always in excess compared to the wt but UPD MASI is less frequent than CNG (Gandhi *et al.*, 2009).

#### 1.5.9 Germline genetics: Susceptibility Loci

Multiple genome wide association studies (GWAS) have uncovered many regions associated with increased risk of lung cancer. One of the most significant associations seen for both smokers and non-smokers is to a region on chromosome 15q25.1 (Amos *et al.*, 2008, Hung *et al.*, 2008, Wang *et al.*, 2008). This region is known to contain several genes (*CHRNA5*, *CHRNA3* and *CHRNA4*) each encoding subunits of the nicotinic acetylcholine receptor, the latter having a known role in nicotine dependence. It is suggested that carcinogens N'-nitrosonornicotine and nitrosamines found in tobacco may facilitate neoplastic transformation by stimulating angiogenesis and tumour growth by interaction with nicotinic acetylcholine receptors (Hung *et al.*, 2008). Through the acetylcholine receptor pathway, cancer cell proliferation, survival, migration, invasion and angiogenesis are promoted (Amos *et al.*, 2008). The gene for proteasome alpha 4 subunit isoform 1 (PSMA4), a component of the ATP- and ubiquitin-dependent non-lysosomal pathway, also resides in this chromosomal region. Unlike the nicotine dependence genes, PSMA4 does not have an obvious part to play in disease etiology and progression (Amos *et al.*, 2008).

Other susceptibility loci that have been reported include a region on chromosome 5p15.33 (McKay *et al.*, 2008, Wang *et al.*, 2008) containing the reverse transcriptase component of telomerase (TERT) (Weir *et al.*, 2007). This is essential for telomerase synthesis and may allow cells to avoid senescence. *CLPTM1L* (cleft lip and palate transmembrane protein 1-like protein), a well conserved gene that induces apoptosis of lung cancer in genotoxic exposures, such as tobacco smoke, as well as enhancing cisplatin-mediated cell death when overexpressed (<http://www.genecards.org>). Both TERT and CLPTM1L are therefore excellent potential targets for therapy.

Chromosome 6p21.33 has also been reported to be associated with risk of disease and within this region of association resides the candidate gene *VEGFA* that may be involved in angiogenesis (Weir *et al.*, 2007, Wang *et al.*, 2008). Other candidate susceptibility genes contained within 6p21.33 include *BAT3* (HLA-B associated transcript 3), which is involved in apoptosis and acetylation of p53, and *MSH5* (MutS Homolog 5), a mismatch repair (MMR) protein.

## 1.6 Treatment and Therapy

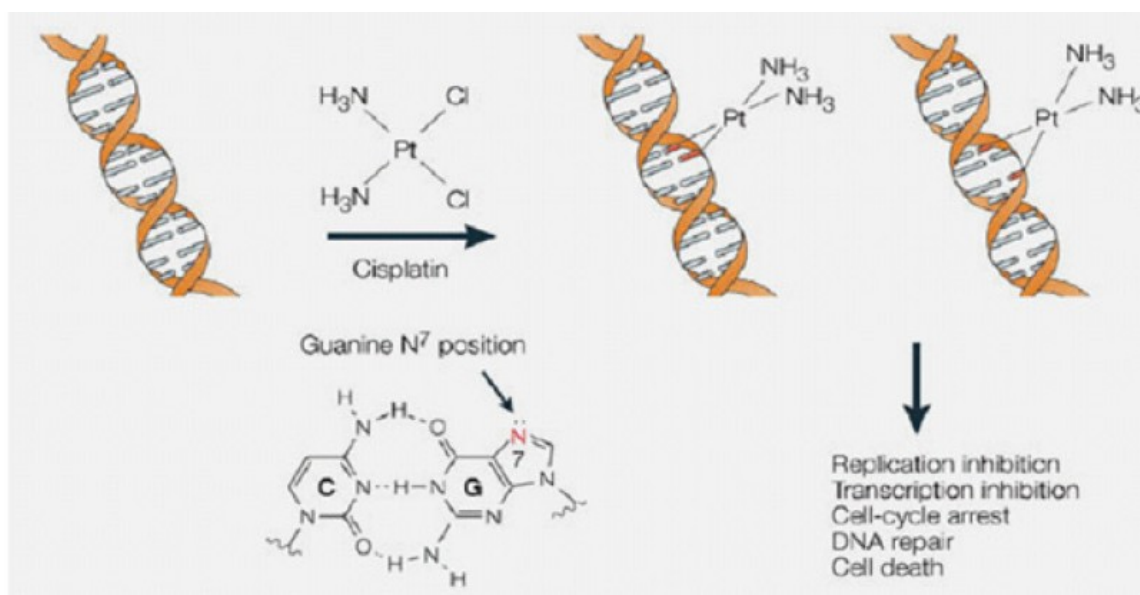
Treatment of lung cancer is usually dependent on cancer type, stage, performance status and co-morbidities, such as Chronic Obstructive Pulmonary Disease (COPD), which may limit the patient's tolerance to chemotherapy.

Due to its early and rapid metastatic capacity, SCLC is regarded as a systemic disease, the cornerstone treatment of which is platinum-based chemotherapy. If diagnosed at an early enough stage, long term survival is achievable by consolidating chemotherapy with radiotherapy. If diagnosed when not “radically” treatable (late stage or extensive stage disease) treatment is palliative, aimed at improving quality of life and survival and is usually through chemotherapy. Although initially exquisitely chemo-sensitive, on relapse SCLC is relatively chemo-refractory and associated with poor survival.

NSCLC treatment is dependent on Stage (see Section 1.2). For Stages I and II, surgery is preferable if the patient is medically fit. If not, radical radiotherapy is an alternative. Unfortunately this is unusual as early diagnosis of the cancer is rare. For Stage II and III tumours, adjuvant chemotherapy may be used after surgery to decrease the chance of relapse. Stage III tumours are usually treated with combined-modality therapy including chemotherapy and radical radiotherapy with-or-without surgery. For patients with Stage IV NSCLC, the main aim of treatment is palliation, to improve survival, control the cancer growth, and alleviation of symptoms through tumour shrinkage. Chemotherapy and small molecule kinase-inhibitor therapy are the main modes of treatment.

Currently, the gold standard of chemotherapy for NSCLC is platinum doublet therapy where a platinum compound; a DNA cross-linking agent such as cisplatin (cis-diamminedichloroplatinum (II)), is combined with another anti-mitotic (such as a nucleoside

analogue [gemcitabine], microtubule inhibitor [paclitaxel], topoisomerase inhibitor [etoposide/ camptothecin]) or a thymidylate synthase inhibitor, such as pemetrexed.



**Figure 1.10** Cisplatin cross-links DNA and in doing so interferes with cell division. Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Drug Discovery (Wang and Lippard, 2005), Copyright 2005.

Cisplatin is one of the most potent anti-cancer therapies and displays significant clinical activity in many solid tumours. Cytotoxicity is generally mediated through apoptosis induction and cell cycle arrest resulting from its interactions with DNA (Figure 1.10). The DNA lesions cisplatin creates are intra-strand DNA adducts and mono-adducts, both repaired by the NER pathway, and inter-strand cross-links. All inhibit and block DNA replication resulting in cell death (Arora *et al.*, 2010, Barr and O'Byrne, 2010).

Cisplatin is one of the most commonly used chemotherapy agents for NSCLC in fit patients, and is generally given to the patient by intravenous injection at 75-80mg/m<sup>2</sup> every 21 days usually in combination with another drug such as gemcitabine, for up to six cycles. Side effects include damage to the kidney and nerves, hearing loss, nausea and vomiting but the latter can be controlled with other drugs.

Targeted therapies have been designed and synthesized based on oncogenes identified to be important in NSCLC carcinogenesis, such as *EGFR*. EGFR inhibitors developed include gefitinib and erlotinib (see Section 1.5.1).

Some NSCLC patients have intrinsically chemo-resistant disease while others become chemo-resistant during the treatment process, with disease that either progresses through or remains stable whilst receiving chemotherapy. In particular the molecular mechanisms underlying platinum-resistance are still largely unknown. Two to 30% of all NSCLC patients are inherently platinum resistant and disease still progresses during early cycles of platinum chemotherapy. It is possible the resistance is mediated by reduced intracellular accumulation of cisplatin, enhanced drug inactivation by metallothioneine and glutathione, increased repair activity of DNA damage, decreased formation of DNA-platinum adducts, and altered expression of oncogenes and regulatory proteins (Adams and Harvey, 2010, Barr and O'Byrne, 2010).

#### 1.6.1 Molecular Mechanisms of Cisplatin Resistance

The mechanisms of cisplatin resistance can be grouped into four categories of alterations; 1) those that involve steps preceding cisplatin binding to DNA, 2) those that directly relate to the binding of cisplatin to the DNA, 3) those concerning the lethal signalling pathway(s) elicited by cisplatin-mediated DNA damage, and 4) those affecting molecular circuitries that do not present obvious links with cisplatin-mediated DNA damage (Galluzzi *et al.*, 2012).

## 1.7 Aims

In order to understand the underlying features of the mechanisms resulting in chemo-resistant lung cancer and disease relapse, genomic approaches will be used, initially in immortalised cell lines and then in patient tissue, with the ultimate hope of identifying potential biomarkers. With this directed approach of controlled exposure to chemotherapy in immortalised cells, it is hoped that, by modelling *in vivo* conditions, important clinically usable biomarkers will be identified. Any potential biomarkers will be investigated further and, ideally, confirmed in patient tissue.

- Establish a cisplatin resistant *in vitro* cell model using the A549 NSCLC cell line.
- Investigate different cisplatin dosing time courses to mirror standard treatment regimes.
- Examine global gene expression changes over the dosing time course using microarrays.
- Perform temporal expression clustering analysis to explore mRNA expression over the cisplatin dosing time course.
- Explore potential co-regulation of genes through network analysis.
- Develop any key findings from the *in vitro* model further by interrogating a set of platinum-treated patient tissue samples.

## **2 General Materials and Methods**

### **2.1 Materials**

#### 2.1.1 Solutions and Reagents

All solutions used in cell culture, MTT assays, and cisplatin dilutions were supplied by Sigma-Aldrich, UK, and prepared under aseptic conditions in a Class 2 microbiological safety cabinet [Biomat Class II] unless otherwise stated. All solutions and reagents used in DNA and RNA extraction and RNA processing for gene expression microarray analysis were provided by the manufacturers stated.

#### **Solutions used in Cell Culture**

##### *Complete Media*

Dulbecco's Modified Eagles Medium (DMEM) was supplemented with 4mM L-Glutamate, heat-inactivated 10% Foetal Calf Serum (FCS), 50000 units of Penicillin and 50mg Streptomycin and kept at 4°C until required. All supplements to DMEM were filter-sterilised first using a 0.2 µm filter cap (Sarstedt, Germany) and a suitably sized syringe.

##### *Trypsin-EDTA (10X)*

5.0g porcine trypsin and 2.0g EDTA were added to 1 litre 0.9% sterile sodium chloride (NaCl). For culture use, this solution was diluted 1:10 with Phosphate Buffered Saline (PBS).

### *Cell Freezing Medium – Dimethyl sulfoxide (DMSO)*

DMSO was combined with heat-inactivated FCS (1:10 dilution) and stored at -20°C until required.

DMSO alone was also a reagent used for the MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) assays (see Section 2.2.2).

### *Cis-diamminedichloroplatinum(II) solution (Cisplatin)*

A ten ml volume of 1mM cisplatin solution was prepared by dissolving three milligrams of cisplatin powder in 0.9% NaCl. The solution was filter sterilized using a 10ml syringe with a 0.2 µm filter cap before immediate use in cell culture experiments.

### *3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) solution*

MTT powder was dissolved in Hanks Balanced Salt Solution (HBSS) to give a 100mg/ml solution. This solution was used to assess cell viability during generation of the cisplatin resistant cell lines (Section 2.1.3).

## **A549 Cell Line**

A549 cells were obtained from American Type Culture Collection (ATCC, <http://www.lgcstandards-atcc.org/>). They are an adherent cell line that is alveolar basal epithelium in morphology with a population doubling time of approximately 22 hours. They were first developed in 1972 through the removal and culturing of cancerous lung tissue in the explant tumour of a 58 year old Caucasian male (Giard *et al.*, 1973). The ATCC standardly conduct characterisation of genotype and karyotype of the parent cells prior to distribution. A549 is a hypotriploid human cell line with a modal chromosome number of 66, occurring in 24% of cells. Cells containing 64 (22%), 65, and 67 chromosomes occur at relatively high frequencies but ATCC have reported the rate of cells with higher ploidies is low (0.4%). There were six markers present in single copies in all cells, including der(6)t(1;6)(q11;q27); ?del(6)(p23); del(11)(q21), del(2)(q11), M4 and M5. Most cells were found to have two X and two Y chromosomes, although one or both Y chromosomes were lost in 40% of 50 cells analysed. Chromosomes N2 and N6 had single copies per cell; and N12 and N17 usually had 4 copies (<http://www.lgcstandards-atcc.org/products/CCL-185>). Further information on the use of A549 cells can be found in Chapter 3 Section 3.1. A range of seeding densities have been used over the course of the experiments detailed in this thesis (Chapter 3 Section 3.2).

## **2.2 Methods**

This section provides detailed information on the laboratory procedures carried out for this thesis. Unless otherwise stated, all practical work was carried out by the author and all centrifugations were carried out using a benchtop microfuge (Eppendorf 5424).

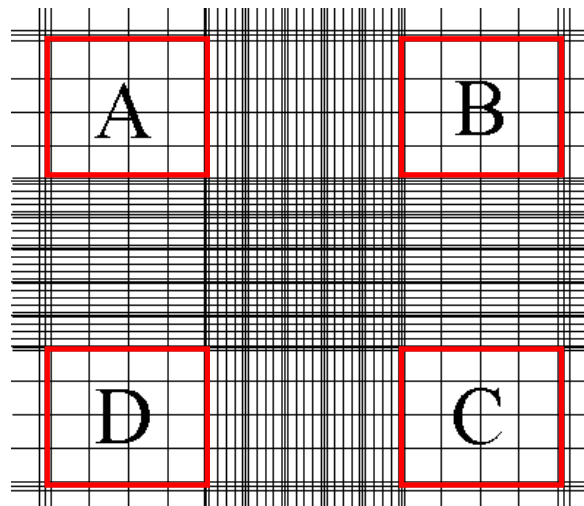
### 2.2.1 Cell Culture of A549 Cell Lines

All cell culture work was conducted in a Class 2 microbiological safety cabinet [Biomat Class II] within a designated cell lines tissue culture laboratory using standard aseptic technique. Unless otherwise stated, all incubations took place in a RS Biotech Galaxy S incubator at 37°C with 5% carbon dioxide (CO<sub>2</sub>)

The frozen aliquot of A549 cells were revived by rapid thawing in a 37°C water bath after which they were then added to 17ml of pre-warmed (37°C) complete media in a 75cm<sup>2</sup> flat bottomed flask [Falcon]. Cells were allowed to revive further by incubating them for three hours during which time they started to adhere to the flask surface. After this, the cells were detached and reseeded in order to remove the dimethyl sulfoxide (DMSO) present in the cryoprotectant medium as it has an element of toxicity. The cells were detached using 10ml of 1X Trypsin-EDTA followed by an incubation of 5 minutes at 37°C after which the trypsin was neutralised by addition of an equal volume of pre-warmed complete media. Cells were then harvested by centrifuging the culture medium (containing the cells) for 4 minutes at 500 x g [Heraeus®, Labofuge® 400R Centrifuge, Thermo Scientific]. The resulting supernatant was removed by suction and the pellet re-suspended in ten ml of complete media, and the required quantity of cells seeded in a 150cm<sup>2</sup> flat bottomed tissue culture flask containing 20ml of complete media.

During the period of cell culture, media was changed every three days or, if the cell density was close to 85% confluence, the cells were passaged. To passage the cells, cells were once more detached from the tissue culture flask surface and then re-seeded at a lower density in a fresh 150cm<sup>2</sup> tissue culture flask. Media was removed using a sterile pipette and cells then washed with 15ml of Phosphate Buffered Saline (PBS) which was then discarded. 10ml of 1X Trypsin-EDTA was added to the flask and then incubated for a minimum of 5 minutes, or

until the cells were detached, at 37°C under CO<sub>2</sub>. Ten ml of complete media was added to the flask to neutralise the trypsin, and cells were then recovered by centrifugation (as detailed above). The resultant cell pellet was re-suspended in 10ml of media and 10µl of this re-suspended cell solution was combined with 90µl of Trypan Blue to allow cell viability counting. 10µl of this mix was applied to a haemocytometer and the viable cells counted under the light microscope. Uncoloured, viable cells in all four large squares (highlighted by red boxes, Figure 2.1) of the haemocytometer were counted, summed, and averaged. The total number of cells calculated using the formula – Cells/ml = average cell count per large square x 10<sup>5</sup>.



**Figure 2.1** Diagram showing the haemocytometer layout under a light microscope. To calculate the number of cells present in a solution, the average cell count per square ( $[\text{number of cells in square A} + \text{number of cells in square B} + \text{number of cells in square C} + \text{number of cells in square D}] \div 4$ ) was multiplied by 10<sup>5</sup> (<http://home.cc.umanitoba.ca/~adam/lab/Haemocytometer.htm>).

A suitable number of cells were taken from this volume and used to seed the required plate or flask. Depending on the use of the cells for the next procedure, this varied from 100,000 to two million cells.

For long-term cryopreservation of the cell lines, a number of cell aliquots were prepared and stored frozen under liquid nitrogen. Cells were allowed to reach 85% confluence before being pelleted as described above for passaging. They were then re-suspended in a sterile solution of 10% DMSO and 90% FCS. One ml aliquots were dispensed into one ml cryotubes that were then placed within a polystyrene box at -80°C for 24 to 48 hours to allow a process of slow freezing to minimise the formation of ice crystals that can disrupt cell integrity. After incubation at -80°C, the cell aliquots were transferred to liquid nitrogen for long term storage.

### 2.2.2 Generation of Cisplatin-Resistant Cell Lines and Round Zero: IC<sub>25</sub>/IC<sub>50</sub> Calculation

An overview of the experimental design for cisplatin dosing of cells is shown in Figure 2.2.

Chemotherapy naive A549 cells were split and seeded into 96-well plates at a density of  $2.5 \times 10^3$  cells per well and incubated overnight. Taking the sterilised 1mM cisplatin (*cis*-diamminedichloroplatinum (II)) solution (see Section 2.1.1) a range of cisplatin concentrations (0.01µM to 100µM) in complete media were prepared by serial dilution. The cells in the 96-well plate were treated in triplicate for each of the concentrations and incubated for 72 hours. Complete media only and cells with complete media (labelled M and M+C controls respectively) were also incubated for the same period of time.

Immediately after incubation, cell viability was measured using the MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) assay. Media with or without cisplatin was aspirated off the cells at the end of the 72 hours incubation. One hundred mg/ml MTT dissolved in HBSS was added to each well and incubated for approximately 45 minutes. During this time, any viable cells present reduce the yellow coloured tetrazolium salt to formazole (purple coloured). Consequently the amount of formazole produced equates

directly to the number of viable cells present. MTT was then removed from the plate and DMSO was used to dissolve the formazole to enable absorbance readings to be taken at 550nm and 595nm. The absorbance readings were then used to calculate the percentage survival compared to M+C controls. This was achieved by using the mean absorbance readings from all the M+C controls on the plate (n=3) as 100% survival for that plate and calculating survival of the other concentrations accordingly. IC<sub>50</sub> values for dose response were calculated using GraphPad Prism version 5.03 for Windows (GraphPad Software, San Diego California USA, <http://www.graphpad.com>). For IC<sub>25</sub> values, the IC<sub>50</sub> value was simply halved. Once calculated, this value was used for the cisplatin dosage of the cells for the subsequent round (Round 1).

Chemo-naive cells were seeded at a density of  $2 \times 10^6$  cells (approximately 70% confluency) for IC<sub>50</sub> 2 week recovery/3 week recovery/4 week recovery or  $0.75 \times 10^6$  well plates cells (approximately 50% confluency) for IC<sub>25</sub> 1 week recovery/IC<sub>50</sub> 2wk recovery. Cells were allowed to adhere overnight. The following day, cells were treated with the correct IC value for 72 hours to generate Round 1 cells, after which, cells were washed with PBS and replenished with complete media. For seeding densities of  $2 \times 10^6$  cells, splitting of cells was performed immediately post drug removal.

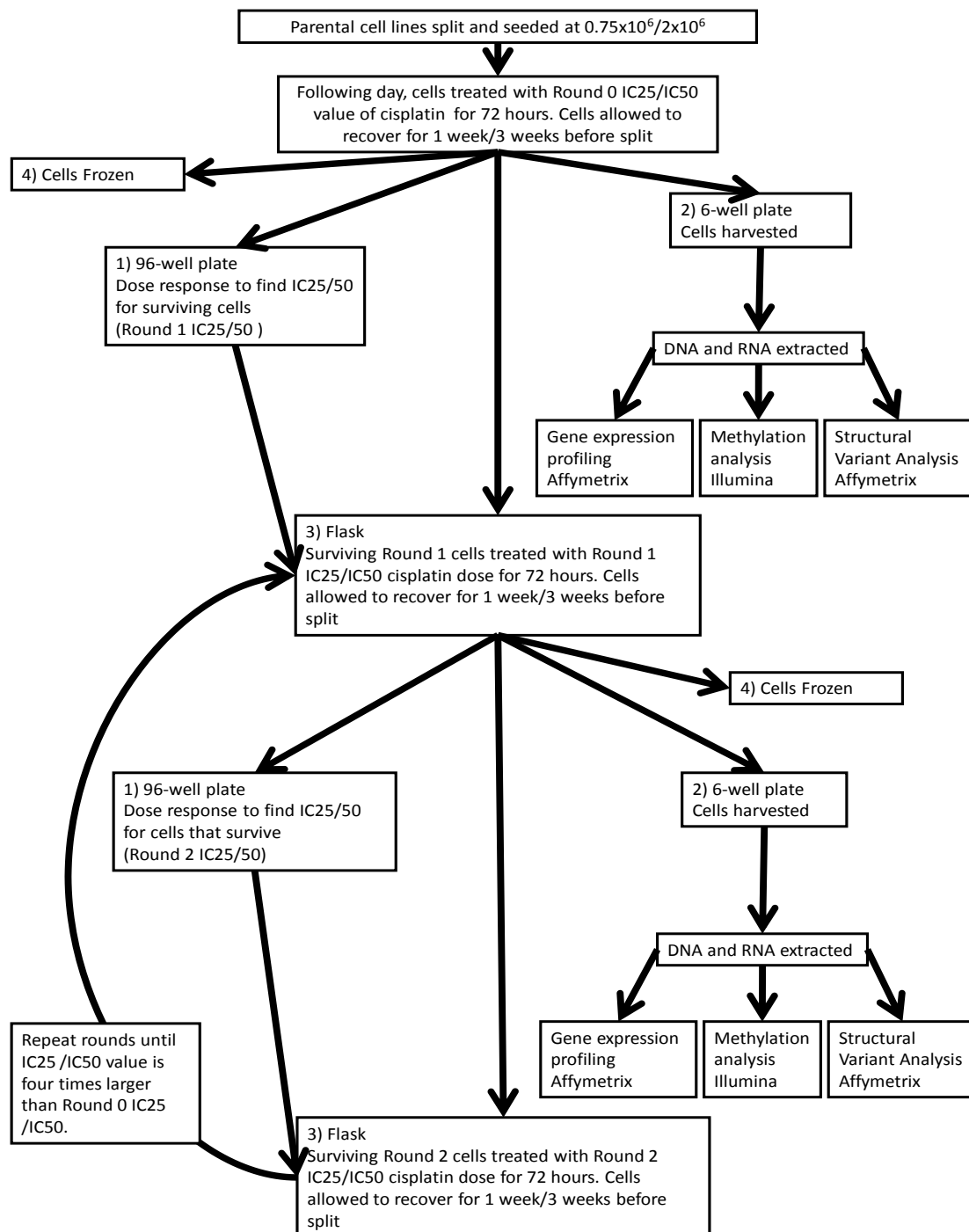
Cells were allowed to recover in complete media, with replenishment with fresh media every three or four days. Cells were split when  $\geq 85\%$  confluency was attained. Recovery periods were maintained for all subsequent rounds of cisplatin dosing and after each recovery period cells were: **1)** split into 96-well plates for Round 1 IC<sub>25</sub>/50 calculations by cisplatin dosing **2)** seeded into 6-well culture plates (at  $7.5 \times 10^4$  cells per well), and cultured until  $>80\%$  confluency for DNA and RNA extraction **3)** transferred to a fresh flask of media for continuation of the round and future splits for drug treatment and **4)** transferred into a fresh

flask for subsequent freezing of an aliquot of cell for long term storage in liquid nitrogen in a cryovial.

The cells for continuation of the round were allowed to recover, reseeded at the correct density and were subject to the next IC<sub>25</sub>/50 cisplatin concentration, calculated from the 96-well plate, for 72 hours (Figure 2.2). This generated the next round (Round 2) of cells. These successive rounds of cisplatin dosing, IC<sub>25</sub>/50 values calculation and DNA and RNA extraction continued until the IC<sub>25</sub>/50 values had increased five-fold. DNA was extracted alongside RNA to provide the option for further and future experiments, such as sequencing analyses across the time course (see Chapter 7 Section 7.1.4).

The cells from the 6-well culture plate were subject to lysis and homogenisation. Six hundred  $\mu$ l of RLT Buffer [Qiagen, UK] containing 1%  $\beta$ -mercaptoethanol was added to each well and incubated for 5 minutes. To maximise lysis, a 200 $\mu$ l pipette tip was used to scratch the bottom of each well, before the solution of buffer and cell lysate was transferred to an autoclaved eppendorf tube and stored at -80°C prior to subsequent RNA extraction (Section 2.1.4.1).

**Figure 2.2** Flow diagram outlining repeated dosing of A549 cells with cisplatin. Figure shows the various time points for dose response and nucleic acid extraction. **1)** 96-well plates used for concentration titrations and subsequent calculation of IC<sub>25</sub>/50 value using the MTT assay **2)** DNA and RNA are harvested for future high throughput genomic investigations. **3)** The same offspring cells subjected to the next round of cisplatin dosing. **4)** Freezing of cells at each round for long term storage in liquid nitrogen.



### 2.2.3 Global Gene Expression

#### *2.2.3.1 RNA Extraction from Cultured Cells*

Before starting any work involving RNA, all surfaces were wiped with 70% ethanol followed by RNase Zap [Sigma]. RNA Dedicated pipettes and pipette filter tips were used at all times to avoid contamination or RNA degradation.

The lysed cells (Section 2.1.3) were carefully thawed on ice and then total RNA extracted according to manufacturer's protocols using the RNeasy Mini Kit [Qiagen]. QiaShredder columns [Qiagen] were used to homogenise and shear the cells by applying the RLT buffer/cell mix to the column and centrifuging at 16,000 x g for 2 minutes. An equal volume, 600µl, of 70% ethanol was added to the QIAshredded lysates followed by mixing by inversion. The samples were successively passed through the column twice to bind the RNA by centrifugation (8,000 x g for 30 seconds). The flow-through was reapplied and spun under the same conditions to increase yield. The spin columns were washed with 600µl of buffer RW1 twice (8,000 x g for 30 seconds) and, again, to optimise yield, each flow-through was re-applied.

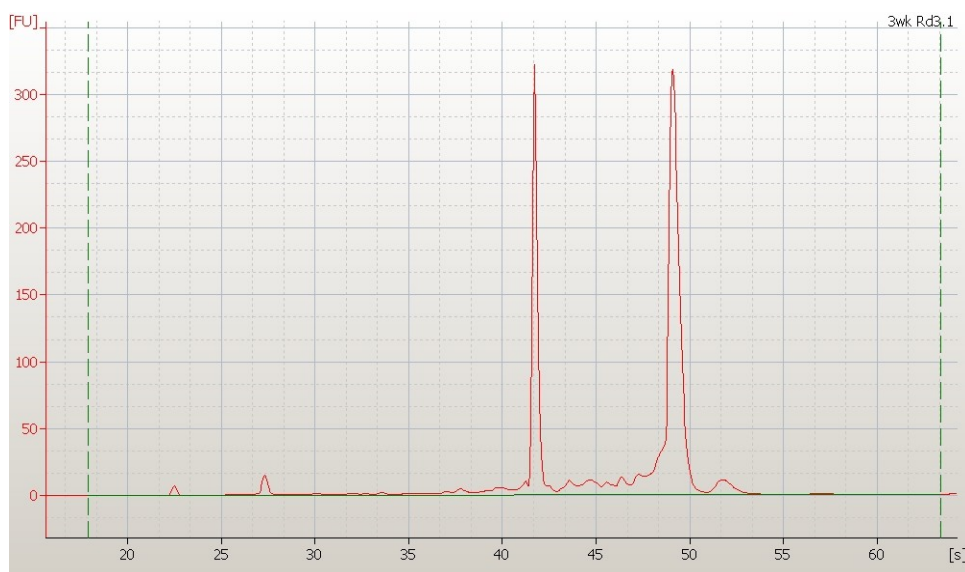
Any DNA contamination in the sample was degraded by incubation with DNase-1. A stock solution of DNase-1 was prepared by dissolving solid DNase (1500 Kunitz units) in 550µl of RNase-free water. For each set of extractions, a solution was prepared using 10µl of this DNase-1 stock solution added to 70µl of RDD buffer, per sample extracted. 80µl of this DNase-1/RDD solution was placed in the centre of each spin column membrane before being incubated for 15 minutes at room temperature (approximately 20°C) on the bench. A further wash of 500µl buffer RW1 was performed with five minutes incubation at the same temperature. The spin column was centrifuged for 30 seconds at 8,000 x g before two washes

with 500µl of RPE, the first for 30 seconds at 8,000 x g and the second for four minutes at 13,000 x g to allow the silica gel membrane to dry fully. RNA was eluted in 30µl RNase-free, Diethylpyrocarbonate (DEPC)-free water [Ambion, Invitrogen], with an incubation of five minutes at room temperature before three minutes at 13,000 x g. The flow-through was reapplied, incubated at room temperature for a further five minutes before another spin of eight minutes at 13,000 x g. This resulted in an increase in overall yield of approximately 30%.

The quantity and quality of the extracted RNAs were measured using a Thermo Scientific NanoDrop® ND-1000 UV-Vis spectrophotometer and the Agilent 2100 Bioanalyzer respectively. Microarray experiments require high quality RNA to obtain accurate, reliable results and an idea of quantity is needed for downstream procedures. A one in three dilution (using RNase- and DEPC-free water) was prepared using a 1µl aliquot of the RNA sample and 1µl of this dilution was used on the NanoDrop® to calculate the concentration in ng/µl, taking into account the dilution factor. The overall yield obtained was calculated using this concentration multiplied by the volume eluted from the spin column.

The Agilent Bioanalyzer chip combines micro-fluidics, capillary electrophoresis and a fluorescent dye (that binds to the nucleic acid) enabling one to access the quality and integrity of an RNA sample. An RNA ladder and 1µl of the one in three RNA dilutions, used for the NanoDrop® concentration calculations, were loaded onto the chip. The sizes and masses of the samples are determined as the fluorescence dye-RNA molecules complex pass through the micro-channels. Analysis of the resultant electropherogram calculates the area under the 18S and 28S ribosomal RNA peaks to give a ratio that indicates the integrity and quality of the RNA, as well as providing an RNA Integrity Number (RIN). Only RNAs that had a

stringent  $RIN \geq 8$  underwent global gene expression. The electropherogram (Figure 2.3) can also be visually assessed for sample integrity/degradation.



**Figure 2.3** Example of an RNA electropherogram from the Agilent Bioanalyzer showing the two distinct peaks on the fluorescence versus time in seconds trace. The 18S peak is located at approximately 42 seconds while the 28S peak sits at approximate 49 seconds. No degradation is visible as shown by the lack of degradation products that would blur the distinction between these two peaks.

#### 2.2.3.2 RNA Preparation and Microarray Hybridisation for Global Gene Expression

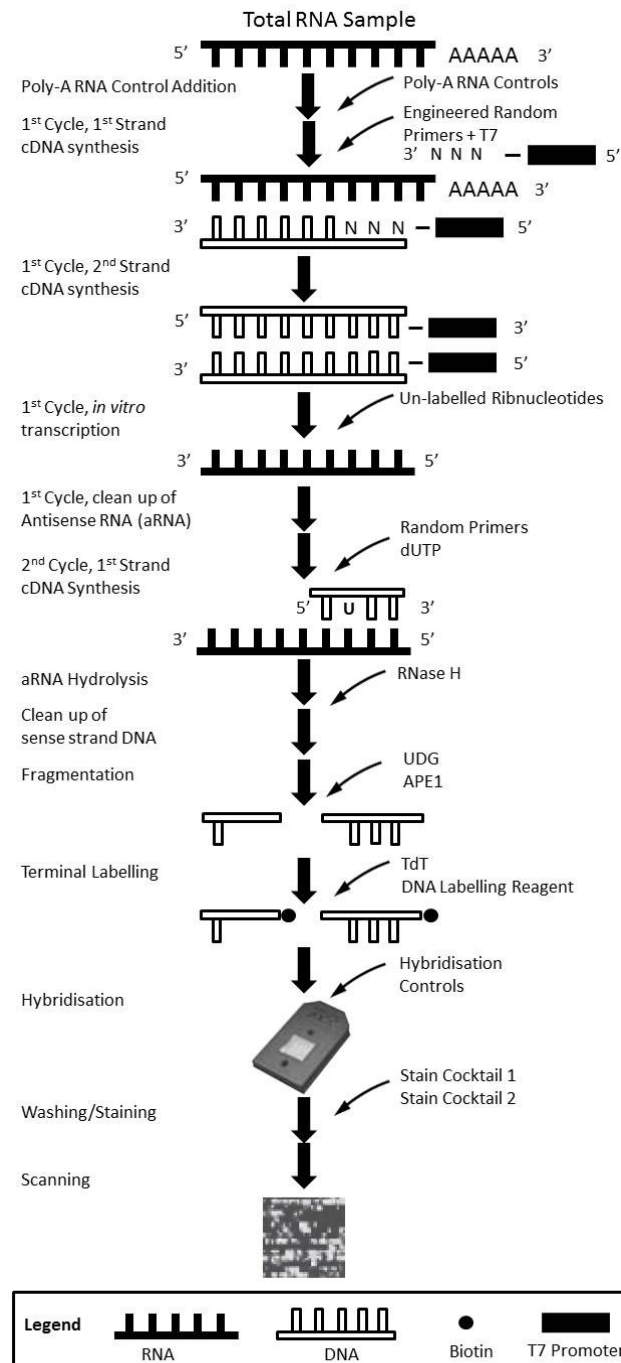
##### *Analysis*

All incubations were carried out on a MJ Tetrad Thermocycler [Genetic Research Instrumentation (GRI)] to ensure uniform incubation temperatures. Samples were processed to fragmentation in batches of four or eight and from labelling to chip hybridisation in batches of 16 or 24, depending on the format of the array plate used.

Each extracted sample RNA processed goes through cDNA synthesis of two strands, the second of which is subject to *in vitro* transcription (IVT) to generate anti-sense RNA overnight. Random primers are annealed to this anti-sense RNA, once it has been purified, to

generate a second cycle cDNA. The anti-sense RNA is then removed by hydrolysis and the remaining cDNA strand is purified ahead of fragmentation into small lengths. This is then subject to labelling, hybridisation to the microarray and subsequent washing and staining.

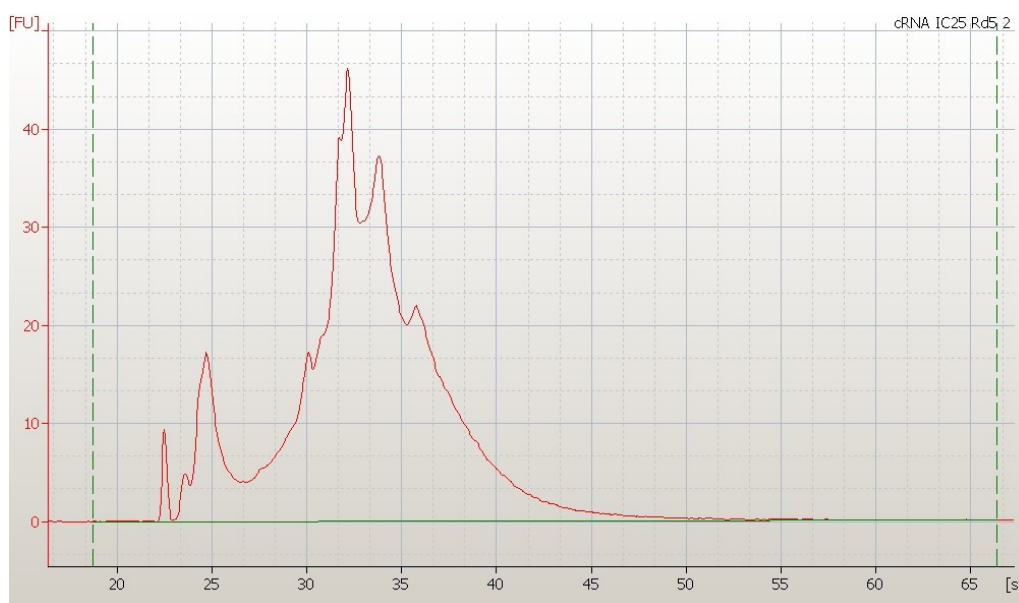
Figure 2.4 shows the schematic of the basic protocol.



**Figure 2.4** Outline of the RNA Processing Protocol, adapted and redrawn from the Affymetrix GeneChip WT Terminal Labeling and Hybridization User Manual.

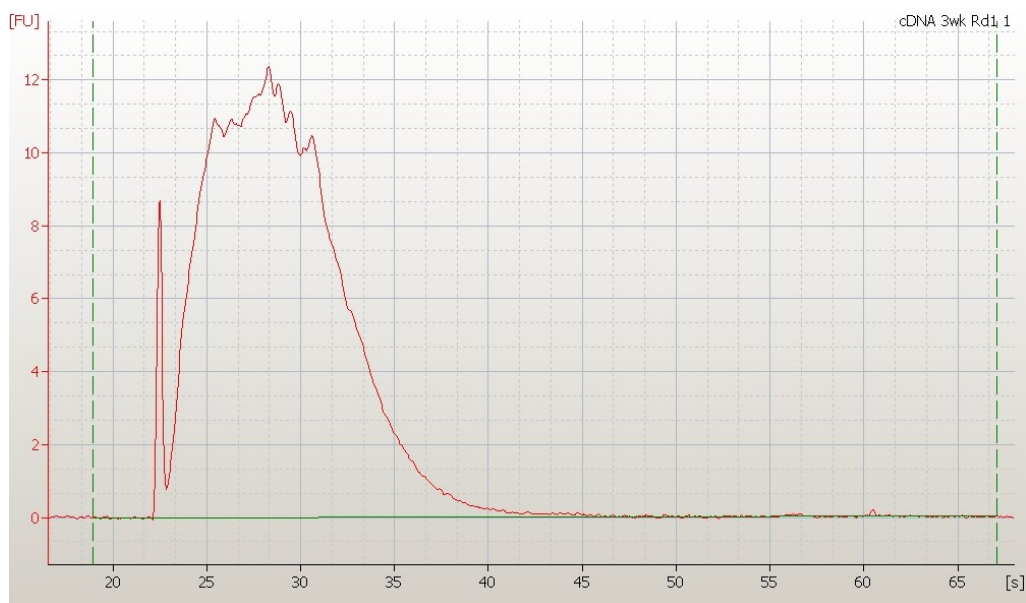
Two hundred nanograms of total RNA was used to synthesise double-stranded cDNA using the Ambion® Whole Transcript (WT) Expression Kit [Applied Biosystems, USA] according to the manufacturer's protocols. This kit is optimized for use with human Affymetrix GeneChip® Arrays ([http://www.ambion.com/techlib/prot/fm\\_4411973.pdf](http://www.ambion.com/techlib/prot/fm_4411973.pdf) and [http://media.affymetrix.com:80/support/downloads/manuals/wt\\_term\\_label\\_ambion\\_user\\_manual.pdf](http://media.affymetrix.com:80/support/downloads/manuals/wt_term_label_ambion_user_manual.pdf)). A Poly-A RNA control was added to the target total RNA sample to provide an exogenous positive control to monitor the entire target labelling process. Five µl of this RNA was subject to first strand cDNA synthesis through a reverse transcription procedure, where the RNA is primed with primers containing a T7 promoter sequence, with incubation for one hour at 25°C followed by a further hour at 42°C. DNA polymerase and RNase H were then used to simultaneously degrade the RNA and synthesise the second cDNA strand by incubating for one hour at 16°C followed by 10 minutes at 62°C. IVT was performed using T7 RNA polymerase, overnight to produce and amplify the anti-sense cRNA. This required 16 hours incubation at 40°C.

The anti-sense cRNA was then purified of enzymes, salts, inorganic phosphates and unincorporated nucleotides to improve its stability, and the yield calculated using the Nanodrop® with size distribution confirmed to yield fragments of the correct size with the majority between approximately 800-1400 nucleotides, using the Agilent 2100 Bioanalyzer (Figure 2.5).



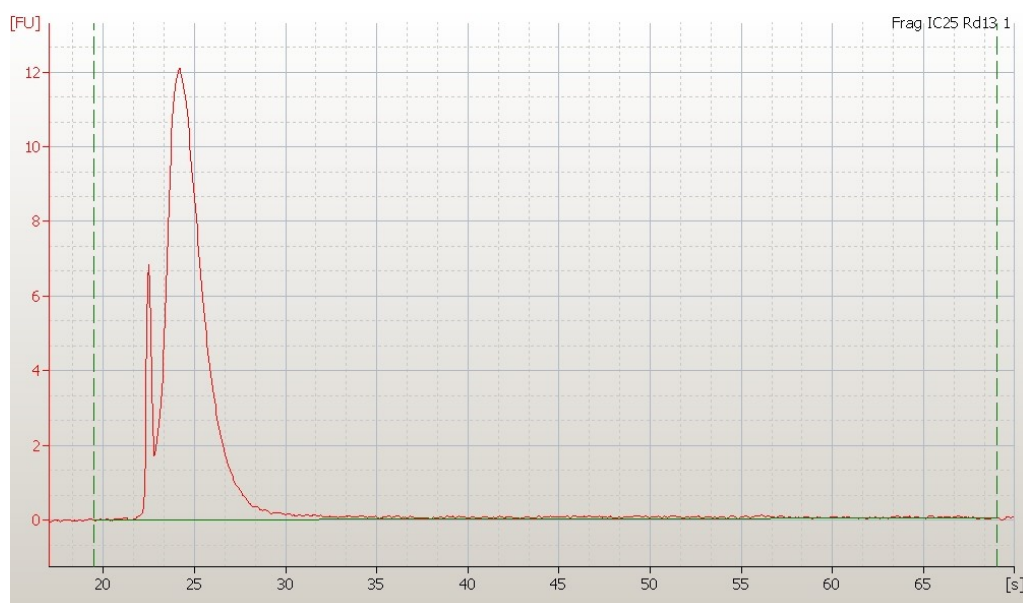
**Figure 2.5** Example of a cRNA electropherogram. It can be seen that the two defined peaks seen in the total RNA sample (**Figure 2.3**) at 42 and 49 seconds no longer appear. Instead, two main less distinct peaks can be seen around 24 seconds and between 28-40 seconds, which is as expected for unfragmented cRNA. Most of these fragments in this second peak will be between 800-1400 nucleotides in length. The presence of spikes within this second peak between 28-40 seconds is due to the enrichment of certain T7 RNA transcript lengths generated in the IVT reaction.

Sense-strand cDNA was synthesised by reverse transcription of cRNA (10µg) using random primers. Two µl of the primers were annealed to the cRNA with incubation for five minutes at 70°C followed by five minutes at 25°C. The reverse transcription involves incubation with the reverse transcription enzymes for ten minutes at 25°C, followed by 42°C for 90 minutes then 70°C for ten minutes. RNaseH, a non-specific endonuclease, degrades the cRNA template in an incubation for 45 minutes at 37°C followed by five minutes at 95°C, to leave single-stranded cDNA, which was then purified as detailed for the anti-sense cRNA, yield calculated using the Nanodrop® and size distribution confirmed to be under 1000 nucleotides for the majority of fragments, using the Bioanalyzer (Figure 2.6).



**Figure 2.6** Example of cDNA electropherogram from the Agilent Bioanalyzer. After the initial marker peak, approximately 22 seconds on the horizontal time axis, there is a clear single curve ranging from approximately 25-35 seconds, indicating the reduction in size of the cDNA from the cRNA (**Figure 2.5**). The majority now are smaller than 1000 nucleotides. This one clear round peak also shows the absence of the spikes observed due to the T7 RNA transcripts.

The Affymetrix GeneChip® WT Terminal Labelling Kit was used to fragment 5.5µg of the prepared single-stranded cDNA. This was performed by incubating this cDNA with 10X cDNA fragmentation buffer, 10U/µl uracil-DNA glycosylase (UDG) and 1,000U/µl apurinic/aprimidinic endonuclease 1 (APE1) for one hour at 37°C followed by two minutes at 93°C. The fragmented samples were once more analysed for size using the Bioanalyzer to ensure all transcripts after this reaction were short nucleotide fragments of between 40-70 nucleotides (Figure 2.7).



**Figure 2.7** Example of a fragmented cDNA electropherogram from the Agilent Bioanalyzer showing all RNAs present in the fragmented sample are of small size between 40-70 nucleotides, as shown by their short run time, between 21s and 27s on the horizontal time axis.

The fragmented single-stranded cDNA samples were stored at  $-20^{\circ}\text{C}$  until 16 or 24 had been prepared after which labelling by terminal deoxynucleotidyl transferase (TdT) using the Affymetrix proprietary DNA Labelling Reagent was performed by incubating at  $37^{\circ}\text{C}$  for one hour, followed by ten minutes at  $70^{\circ}\text{C}$ . The hybridisation cocktails and hybridisation efficiency controls were then added: 3nM of Control Oligonucleotide B2 and 20X Eukaryotic Hybridisation Controls (BioB, BioC, BioD, Cre). The hybridisation cocktails were denatured at  $95^{\circ}\text{C}$  for five minutes followed by  $45^{\circ}\text{C}$  for 5 minutes before being placed in the hybridisation plate.

The GeneTitan® Instrument [Affymetrix] was prepared for the Gene 1.1 Sense Target (ST) Array Plates by preparing the wash buffers A and B and the stains 1, 2 and 3 (Table 2.1). The stain trays and covers used to hold the stains were static neutralised by using an anti-static gun that emits a stream of positive ionised air over the objects surface as the trigger is pulled

and a stream of negative air ions as the trigger is released. The required stain solutions and array holding buffer were placed in the necessary wells of the static-neutral trays and loaded into the GeneTitan® according to the manufacturer's protocols, along with the array plate, the Scan Tray and the hybridisation plate containing the denatured hybridisation cocktail samples. Sample names and locations on the arrays were entered into the control portal to allow tracking and subsequent labelling of the data generated. The use of the GeneTitan® enables multiple samples to be run in parallel, hence samples being run in batches of 16 or 24. Five runs in total were conducted to ensure complete coverage of the two chosen cisplatin dosing regimes; three runs used 16 array plates and two runs used 24 array plates. Each triplicate sample, prepared at the rounds of dosing during the treatment time course, was run for validation purposes.

| Reagent              | Amount Required for One Array Plate | Number of Plates that can be Processed using the GeneTitan Hybridization, Wash and Stain Kit for WT Array Plates (P/N 901622) |           |
|----------------------|-------------------------------------|-------------------------------------------------------------------------------------------------------------------------------|-----------|
|                      |                                     | 16-Format                                                                                                                     | 24-Format |
| Wash A               | 920 mL                              | 1                                                                                                                             | 1         |
| Wash B               | 300 mL                              | 1                                                                                                                             | 1         |
| Stain 1 and 3        | 105 µL/well                         | 6                                                                                                                             | 4         |
| Stain 2              | 105 µL/well                         | 6                                                                                                                             | 4         |
| Array Holding Buffer | 150 µL/well                         | 6                                                                                                                             | 4         |

**Table 2.1** Volumes required to process Gene 1.1 ST Array plates per run

At all stages in the RNA processing pipeline, from extraction to microarray running, experiments were performed in batches. These were carefully recorded as any substantial batch effects can be corrected for during subsequent analyses. An example of a common batch effect occurs when samples that need to be compared are run across several array

plates; the fluorescence intensities between the array plates may not be the same, purely due to the difference in experimental variables, such as the array plate itself, and so batch correction is used to allow these comparisons to take place.

#### *2.2.3.3 Global Gene Expression Microarray Quality Control and Normalisation*

Raw intensity measurements of all probe sets were corrected for background fluorescence, normalised and converted into expression measurements using Affymetrix Power Tools (APT). The CEL files of microarrays were subject to probeset summarisation through robust multi-array averaging (RMA). Post processing was conducted in R Statistics v2.15 with associated packages, “Affy” and “Limma” [both Bioconductor, USA (<http://www.bioconductor.org>)] (Chapter 4 Section 4.2.2).

#### *2.2.3.4 Differential Expression Analysis*

Differential expression analysis was conducted in R with associated packages “Affy” and “Limma” (Linear Models for Microarray Data) [Bioconductor (<http://www.bioconductor.org>)]. Across the time courses, each transcript cluster’s RMA-normalised expression at each and every time point was contrasted with every other in the dosing regime, i.e. in the IC50 Dose 3 Week Recovery regime time point 0 was contrasted with time points 1 to 10, time point 1 was contrasted with time points 2 to 10, time point 2 was contrasted with time points 3 to 10 etc. until all points and genes were compared. This generated a T statistic for each gene and from all the T statistics, a moderated F statistic was generated taking into account all possible comparisons and allowing examination for any

significant change across the whole time course, giving a single test for each transcript cluster on the Gene 1.1 ST array. By using the F-distribution, the F-statistics generate  $p$ -values.

The  $p$ -value for a particular gene is equal to the probability that the expression levels over the time course come from the same distribution therefore if it is small it infers significant differential expression. The overlap of the statistical distributions is important and will depend on the variances. Multiple testing and false positives may cause problems with these data sets as thousands of tests are performed. To avoid these potential pitfalls, the Benjamini & Hochberg (B-H) correction method was applied to control of the expected false discovery rate (FDR), which sets the probability of a false positive. A threshold FDR-adjusted  $p$ -value of less than or equal to 1E-08 were used for cell culture experiments in order to reduce the number of genes taken through to further analysis to a more manageable number, approximately 1,500 genes (Chapter 4 Section 4.3).

#### *2.2.3.5 Temporal Expression Profiling by Fuzzy C-means Clustering*

For fuzzy  $c$ -means clustering, data that were summarised at the transcript cluster level were imported into R.  $C$ -means clustering was used to group the significantly differentially expressed transcripts into discrete and stable clusters of time series data using the “Mfuzz” package [Bioconductor (<http://www.bioconductor.org>)].  $C$ -means clustering allows each transcript to belong to a particular cluster to a certain degree or weighting; it is either strongly or weakly associated with a cluster, it does not sit within a cluster. This allows smaller changes to be detected than the more stringent and ubiquitous hard  $k$ -means clustering.

The genes in each cluster were analysed using Database for Annotation, Visualisation, and Integrated Discovery (DAVID, <http://david.abcc.ncifcrf.gov>) (Dennis *et al.*, 2003) to provide

Gene Ontology (GO) term enrichment analysis, functional annotation clustering as well as details of the gene families found within the annotation clusters (Chapter 4 Section 4.4).

#### *2.2.3.6 Weighted Gene Co-expression Network Analysis*

The differential gene expression data generated in Chapter 4 Section 4.3 has also been subject to weighted gene co-expression network analysis (WGCNA [Langfelder and Horvath, 2008]) in order to provide insights into potential signalling pathways of interest and suggest genes that may act as hubs. It allows correlation patterns among genes across microarray samples to be described and find modules strongly correlated to traits of interest as well as highly interconnected genes that may be potential hubs. In this work the trait of interest used was the IC<sub>25</sub> or IC<sub>50</sub> value, depending on the dosage used for that particular time course. Genes within each module were analysed using DAVID to provide GO term enrichment analysis, functional annotation clustering as well as details of the gene families found within the annotation clusters, if appropriate (Chapter 5 Section 5.2).

#### 2.2.4 DNA Extraction from Cultured Cells

DNA extractions were performed according to manufacturer's protocols [Qiagen]. DNA was extracted from harvested cells using the QIamp Mini Kit. The harvested cell pellet was subjected to Qiagen protease action and cell lysis using buffer AL before incubation at 56°C for 10 mins. The lysed cells then underwent ethanol precipitation and multiple washes in spin columns using buffers AW1 and AW2 to bind the DNA to the column. Elution was conducted with buffer AE and the eluted DNA was stored at -20°C until use in further

analyses. The quantity of the extracted DNA was measured using the Thermo Scientific NanoDrop®.

### 2.2.5 Lung Cancer Patient Tissue Work

Samples were obtained from two sources for use in this thesis; RESTORE-AIR – Randomised Evaluation of STents to Open REstricted AIRways in patients with centrally placed non-small cell lung cancer, and EQUALITY – Expression QUAntitative trait Loci mapping In Tumours collected at biopsy. Information pertaining to these two studies can be found in Chapter 6 Section 6.2 and Section 6.3 respectively. Results from tissue extraction are described in Chapter 6 Section 6.4. The combination of these two datasets to allow comparisons between platinum treated versus non-platinum treated patients as well as the necessary outlier detection and removal and microarray quality control is detailed in Chapter 6 Section 6.5. Differential expression analyses conducted were similar to those outlined in Section 2.2.3.4 and is defined in Chapter 6 Section 6.6.

#### *2.2.5.1 Collection of Patient Tissue and RNA Extraction*

After obtaining the clinical diagnostic sample from rigid bronchoscopy or biopsy, three sets of research biopsies were taken. Tumour tissue and paired normal carina tissue were biopsied and immediately placed in RNase inactivation buffer (RNAlater [Qiagen]). A second set of tumour and paired normal tissue were snap frozen in liquid nitrogen and subsequently stored at -80°C. A final set of tumour and paired normal tissue were subsequently fixed and processed by routine methods (Formalin-Fixed Paraffin-Embedded [FFPE]) in the hospital

Histology Department. Rigid bronchoscopy can only access proximal tumours and thus, most tumours were either squamous NSCLC (majority) or SCLC.

The extraction of total RNA from patient lung tissue involves the same overall steps as outlined in Section 2.2.3.1, i.e. lysis and homogenisation, Proteinase K digest, total RNA binding, wash and DNase digest, and elution of the RNA.

Each sample was thawed from -80°C storage and weighed to determine its size. Three hundred µl of RLT lysis buffer [QIAGEN], containing 1% β-Mercaptoethanol [Sigma], was added to the tissue before homogenisation with the TissueRuptor. The sample was subjected to several 30s homogenisations up to a maximum of five with the sample being held on ice in between bursts until the lysate was uniformly homogeneous. 90µl of RNase-free water was added to the lysate along with 10µl of Proteinase K solution [QIAGEN], mixed thoroughly before incubation at 55°C for two hours.

After the incubation, the samples were centrifuged at 13,000 x g for three minutes at room temperature. The supernatant (900µl) was transferred into a new microfuge tube and 450µl of 100% ethanol was added and mixed by pipetting. Seven hundred µl of the sample, including any precipitate formed, were placed in an RNeasy mini spin column inside a 2ml collection tube and centrifuged at room temperature for 15s at 8,000 x g. The flow through was discarded, the remainder of the sample added to the RNeasy mini spin column and the centrifugation repeated. 350µl of Buffer RW1 was used to wash the spin column by centrifugation for 15s at 8,000 x g at room temperature. A DNase digest followed as detailed in Section 2.2.3.1. A further wash of the spin column with 350µl of Buffer RW1 followed using the same centrifuge settings and the flow through was discarded. The spin column membrane was further washed with 500µl of Buffer RPE by centrifugation for two minutes at 8,000 x g at room temperature. The spin column was then transferred to a new collection tube

and centrifuged at full speed for one minute to dry the membrane and reduced the possibility of carrying any Buffer RPE over. The bound RNA was eluted from the spin column by the addition of 50µl of RNase-free water to the membrane, followed by centrifugation for one minute at 8,000 x g.

### **3 Generation and Optimisation of a Model of Cisplatin Resistance using A549 Cells**

#### **3.1 Introduction to Cell Culture Models**

Model systems using different types of tissue culture to investigate the biology behind various conditions *in vitro* are well known. There are various types of culture, each with their own advantages and disadvantages (Masters, 2000). A primary culture is created by directly taking tissue from an individual, and then extracting and growing cells from it *in vitro*. It is possible to create a cell line so that, once transferred to a subsequent culture flask, a relatively uniform population of proliferative cells will emerge as a result of selection. They will continue to grow until the end of their natural proliferative lifespan but senescence will occur. Little tissue-specific differentiation occurs during proliferation, although functional tissue may be regenerated, given the correct signals. These normal cells have a limited lifespan in culture so they can only be used over a short period before they senesce.

If dealing with cancerous tissue, the cells can provide a snapshot of the tumour at the time of biopsy and be representative of the cells that divide when the tissue of origin is wounded. In spite of the fact that cancer cells express telomerase and so can elongate their telomeres, potentially delaying the inevitable senescence observed with normal cells. They cannot be cultured easily and potentially still senesce anyway, so transfection with viral gene products is common to extend their lifespan. These products can sequester genes of the cell, such as p53 and Rb, thereby increasing potential passage numbers. The similarities seen between this immortalisation of cell lines and carcinogenesis are evident as the cell lines begin to acquire the hallmarks of cancer (Chapter 1 Section 1.4). Only highly aggressive cancer cells have accumulated enough genetic changes required for unlimited growth *in vitro* (Masters, 2000). These continuous cancer cell lines, such as the A549 cells used in the investigations outlined

in this thesis, need to be grown with commercial tissue culture medium containing FCS, which supplies the source of nutrients such as hormones and growth factors that are needed for the cells to grow and proliferate. Cancer cells from tissue can be extremely difficult to grow in culture. In spite of their apparently limitless growth potential, it is sometimes easier to generate a normal tissue cell line than one from the cancer cells themselves. Invariably, only metastatic cancers can be established as a continuous cell line as they have gathered sufficient number of driver mutations to undergo continued uncontrolled growth.

Although few studies have examined the representativeness of cancer cell lines with their tumours of origin, direct comparisons have been made with respect to genotype and phenotype. These have revealed that cultures retain many properties from the corresponding tumour for long periods, with reports of NSCLC cell lines retaining properties for a median period of over three years (Wistuba *et al.*, 1999). Gene expression studies have revealed consistent similarities between cell lines from the same origin and consistent differences between cell lines of different origins. Even drug sensitivity has been shown to be retained in cisplatin-hypersensitive testicular cancer cells (Walker *et al.*, 1987). Cell line models however tend not to represent the variety and heterogeneity of cancers seen in the clinic. As cell lines require multiple changes to enable growth *in vitro*, late stage, fast growing, poorly differentiated tumours tend provide continuous cell lines (Masters, 2000).

Cell lines do provide a virtually unlimited supply of cells with similar genotype and phenotype. Their use therefore avoids variation between individuals and bypasses ethical and potential cost issues associated with their equivalent *in vivo* experiments. The use of a cell line also allows the investigator more control over variables, such as drug exposure, than would be possible in animal or human systems. As culture systems become more advanced, a

greater proximity and similarity to the organism of interest is possible, as seen with three-dimensional tissue culture.

The established A549 cell line is widely used as an *in vitro* model for type II pulmonary epithelial cell models for drug metabolism and so provides a suitable model system to examine the genetic responses of lung tumours to chemotherapy. Details on A549 cell line genotype and karyotype can be found in Chapter 2 Section 2.1.1. Further important genetic information about generic A549 cells, including potentially cancer-causing mutations (Chapter 1 Section 1.6), are given in Table 3.1 (<http://cancer.sanger.ac.uk/cancergenome/project/cosmic>).

| Sample | Gene                         | AA Mutation                | CDS Mutation                     | Somatic Status                               | Zygosity   | Vaildated |
|--------|------------------------------|----------------------------|----------------------------------|----------------------------------------------|------------|-----------|
| 905949 | <a href="#">CDKN2A</a>       | <a href="#">p.0?</a>       | <a href="#">c.1_471del471</a>    | Reported in another cancer sample as somatic | Homozygous | Verified  |
| 905949 | <a href="#">CDKN2a (p14)</a> | <a href="#">p.0?</a>       | <a href="#">c.1_522del522</a>    | Variant of unknown origin                    | Homozygous | Verified  |
| 905949 | <a href="#">KRAS</a>         | <a href="#">p.G12S</a>     | <a href="#">c.34G&gt;A</a>       | Reported in another cancer sample as somatic | Homozygous | Verified  |
| 905949 | <a href="#">SMARCA4</a>      | <a href="#">p.Q729fs*4</a> | <a href="#">c.2184_2206del23</a> | Variant of unknown origin                    | Homozygous | Verified  |
| 905949 | <a href="#">STK11</a>        | <a href="#">p.Q37*</a>     | <a href="#">c.109C&gt;T</a>      | Reported in another cancer sample as somatic | Homozygous | Verified  |

**Table 3.1** Table indicating the known and verified cancer point mutations in A549 cells (Catalogue of Somatic Mutations in Cancer – Wellcome Trust Sanger Institute). AA mutation = Amino acid mutation, CDS mutation = coding sequence mutation.

A549 cells have been used in the investigations in this thesis due to them being a manipulable system that can be exposed to the conditions under exploration. They are resilient and will grow for several passages and potentially in sub-optimal conditions. They have been used in many studies into various lung conditions including other lung cancer models. Antibiotics were added to the media to enable the long term passage of the cells, as required for

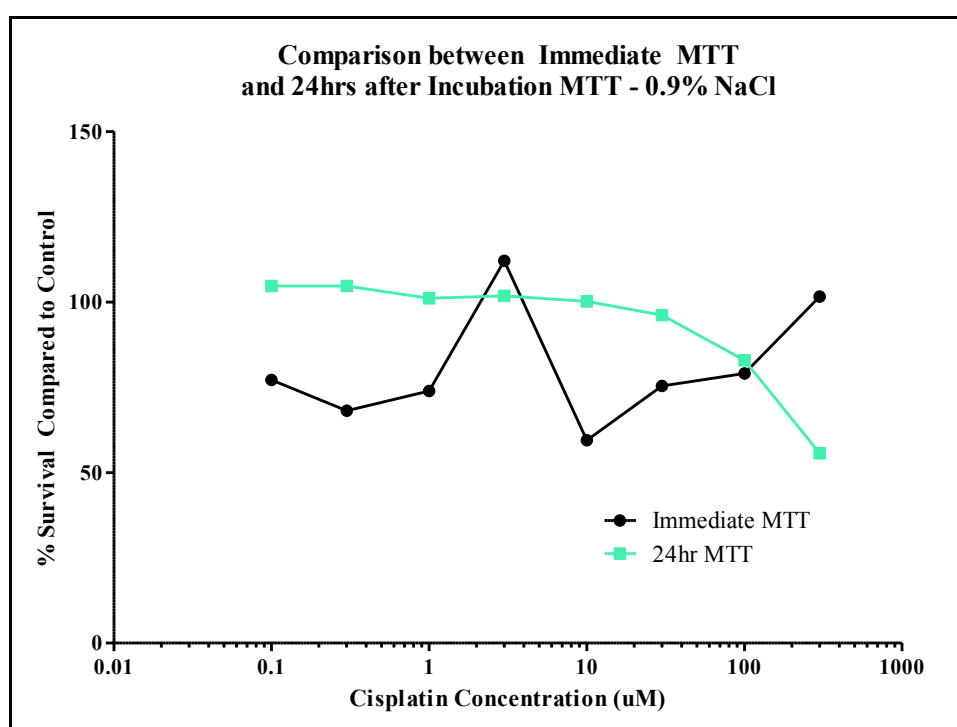
performing repeated measures on the successive generations, without the risk of acquiring infections. While the use of primary cancer cell culture, grown from a biopsy of tumour tissue, would provide a different system in which to characterise the potential response to chemotherapy, A549 cells provide an easy-to-set up, replicable, low-cost model that can survive for several months to provide the platform for the long time courses outlined in this chapter.

### **3.2 Optimisation of Repetitive Cisplatin Dosing and Associated Techniques**

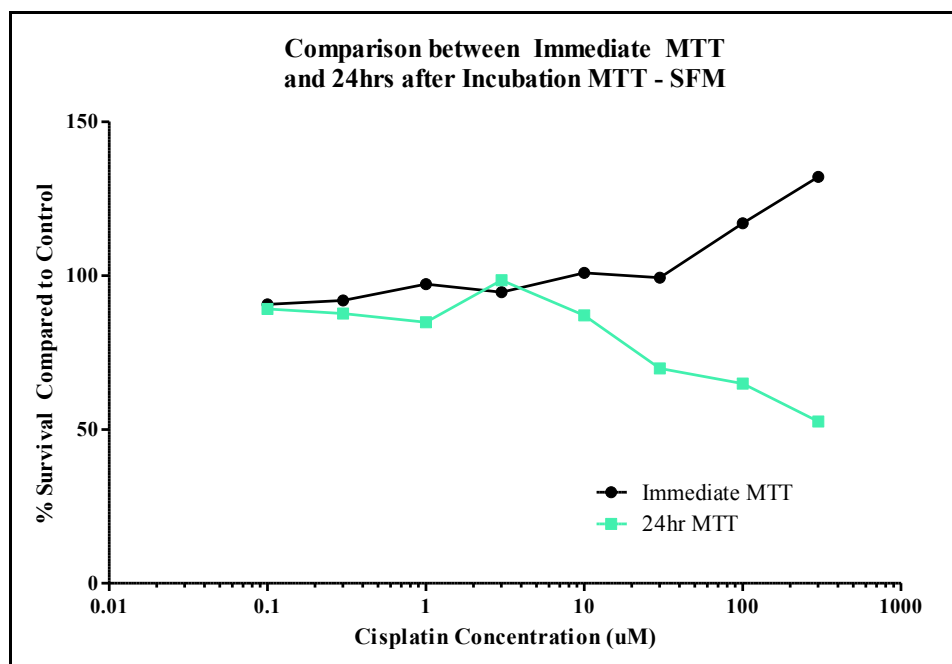
In order to establish the cell model, a number of variables had to be optimised. At the time of cell model initiation, there was little in the published literature on increasing cisplatin resistance in lung cancer cell lines through incremental steps over time. Comparisons were therefore drawn from other cell models; Li *et al.*, (2009) induced cisplatin resistance by subjecting a single cisplatin-sensitive clone of the ovarian cancer cell line A2780 to multiple cycles of increasing doses of cisplatin. Cells were treated to three hour incubations of a range of cisplatin doses from 0.5 to 100  $\mu\text{M}$ , allowed to recover for three days before the next round of cisplatin dosing. MTT assays were used to determine cell viability and subsequent dose response curves. This basic outline of a method was used as the starting point for the experiments carried out.

As per standard protocols within the Molecular Genetics and Genomics Section (NHLI), cells were seeded at a density of  $2.0 \times 10^5$  cells per well, followed by a 24 hour starvation step before drug dosing commenced. For the treatment of cells with 1mM cisplatin dissolved in 0.9% NaCl solution (Chapter 2 Section 2.1.1), different solutions were used, including 0.9%

NaCl solution alone as it would be administered in a clinical setting, or as an addition to serum-free media (SFM). Concentrations of the drug ranging from 0.1 $\mu$ M to 300 $\mu$ M and were performed in triplicate. MTT assays were performed both immediately after drug removal and 24 hours later. Each dose response plate was normalised to the controls (cells with drug vector only) for that particular plate (Figures 3.1 and 3.2). It can be seen that the MTT performed 24 hours after the end of the incubation with cisplatin using either drug vector led to a decrease in cell viability at the higher drug concentrations. This may be indicative of the cisplatin impacting the viability of the cells some time after their incubation.

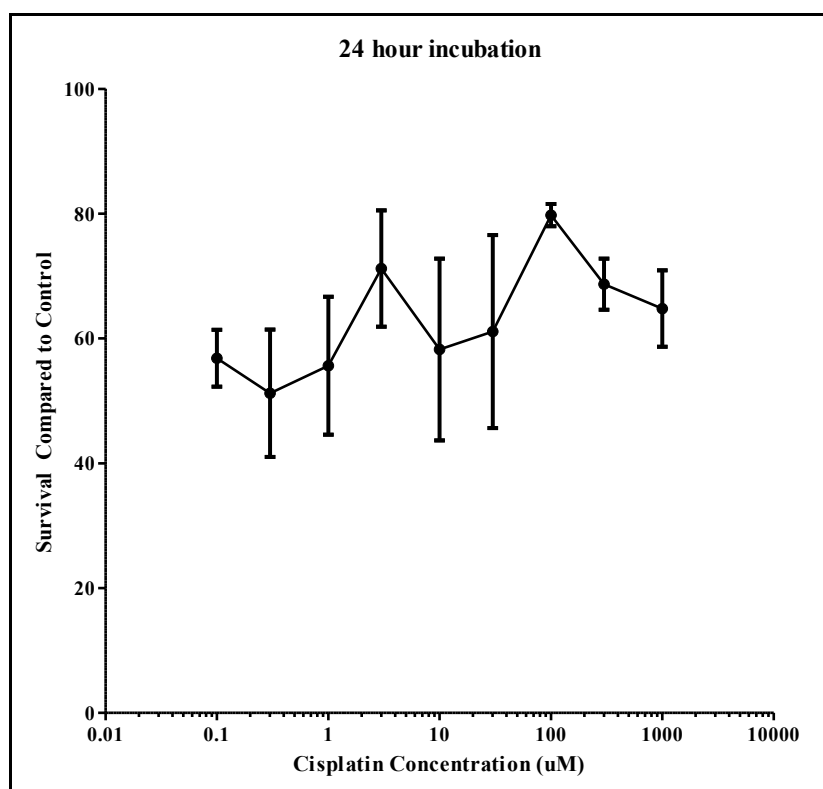


**Figure 3.1** Dose response with cisplatin in 0.9% NaCl solution after a 3 hour incubation for the concentration range 0.1 $\mu$ M to 300 $\mu$ M.



**Figure 3.2** Dose response for cisplatin in serum free media (SFM) after 3 hour incubation over a range of concentrations from 0.1 $\mu$ M to 300 $\mu$ M. Minimal difference was observed between the MTT performed immediately after incubation and 24 hours later for the lower concentrations of cisplatin below 30 $\mu$ M but a divergence in viability seems to occur at the higher concentrations.

When 0.9% NaCl solution was used for drug delivery, problems arose during aspiration with apparently viable cells, both treated and untreated with cisplatin, (purple following MTT application) failing to remain adherent to the bottom of the plate. Using SFM instead of 0.9% NaCl solution for cisplatin delivery appeared to solve both these problems. From Figure 3.3, it can be seen that cells treated with cisplatin using SFM have a range of viability across all drug doses trialled. The desired gradual decrease in viability at the higher cisplatin concentrations, seen to a slight but insufficient extent in the 24 hours MTT data in Figure 3.2, is not evident.



**Figure 3.3** Dose response of cells seeded at  $0.1 \times 10^6$  cells per well with triplicate cisplatin concentrations in SFM incubated for 24 hours. Data shown is mean survival ( $n=3$ ) with standard error of the mean (SEM) error bars. There appears to be little change in cell viability (survival compared to controls) across the cisplatin concentrations as almost all the error bars overlap one another.

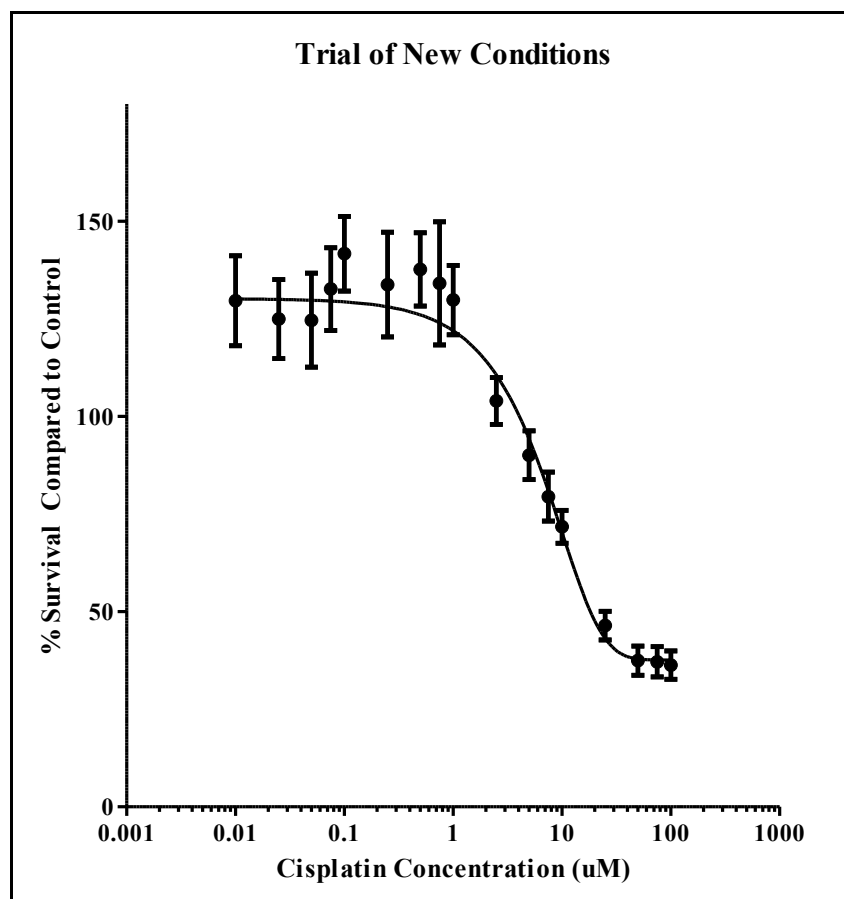
### 3.2.1 Variability due to Seeding Density

Another variable that required optimisation was cell seeding density. It is possible that difference in the initial number of cells present on cisplatin application could vary their subsequent survival; if they are stressed, such as over-confluence, before cisplatin exposure, they may not respond to the drug in the same way. Cell seeding was therefore decreased to 2,500 cells per well to ensure the cells were less than 70% confluency at the time of drug dosing; this also approximately matched the work done by Li *et al.*, (2009). In reducing the seeding density, complete media could also be tested as a drug delivery method, and thus allow the cells being treated with cisplatin to be more representative of the *in vivo*

environment, i.e. fed with the normal nutrients required for growth, without the potential problem of over-confluence.

Cisplatin was dispensed in complete media and untreated cells were maintained in the same complete media over the period of incubation, which was also extended to 72 hours. Cells were seeded the day before drug treatment and unlike previous experiments, no serum-starvation step was performed in an attempt to replicate an *in vivo* situation and to see if this had an impact on cell viability. MTT assays were carried out immediately after drug removal.

Although there was a dramatic improvement in the results obtained through the implementation of these changes, it was still found that control cells failed to survive better than some of the lower cisplatin concentrations (Figure 3.4). This was rectified by adding a one well exclusion zone, filled with PBS, around the edge of the 96-well plates used in cisplatin dosing in order to prevent evaporation of media or cisplatin solutions. In addition to this, an extra media only and Media + Cells (M+C) control was added to the central region of the plate (Figure 3.5).



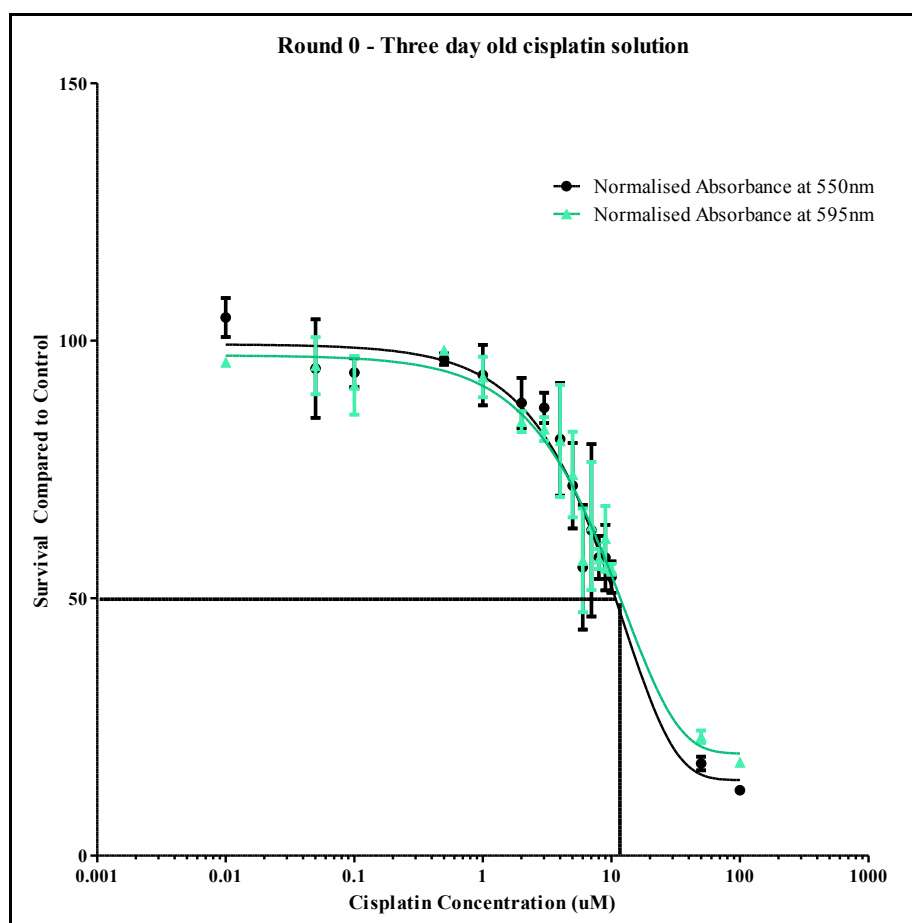
**Figure 3.4** Dose response using new conditions of 2,500 cells per well, 72 hour incubation with cisplatin in complete media with MTT performed immediately after drug removal. Error bars show SEM, where n=6. The percentage survival of cells at the higher concentrations of cisplatin reflects the burden of the drug on replicating cells. The data here could also imply that cells exposed to low concentrations of cisplatin appear to survive better than controls in complete media.



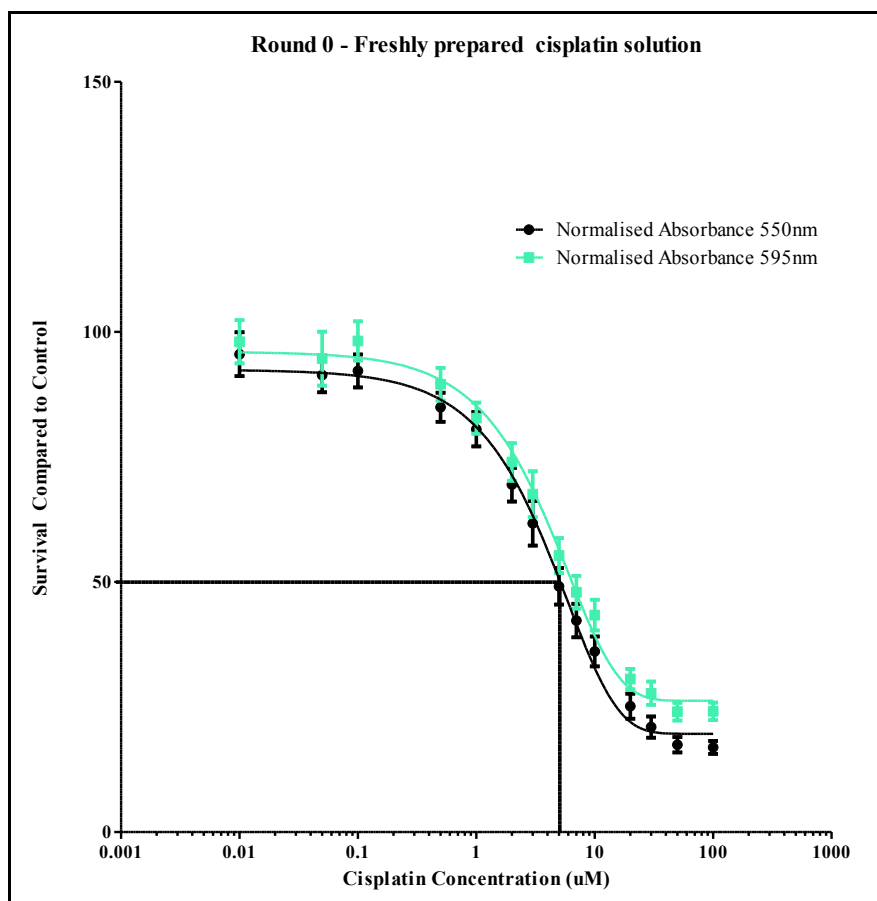
As these conditions yielded such an enhancement on previous attempts, they were used as the basis for all future cisplatin incubations. These changes enabled IC<sub>50</sub> values to be calculated with ease due to the alteration in survival across the cisplatin concentration titration. In spite of this, repeats of cisplatin dosing on chemo-naive A549 cells found there were inconsistencies in IC<sub>50</sub> values obtained. It was therefore decided to investigate the stability of the cisplatin (Section 3.2.2).

### 3.2.2 Cisplatin Stability

Cisplatin is a light sensitive compound and degradation can occur after storage. It was established that, after just three or four days of storage, large differences in IC<sub>50</sub> value were observed. As can be seen from Figures 3.6 and 3.7, Round 0 IC<sub>50</sub> value was found to differ by 7µM between a fresh cisplatin solution and one prepared a few days earlier. Consequently, for all subsequent experiments cisplatin concentrations were made up fresh on the day of treatment.



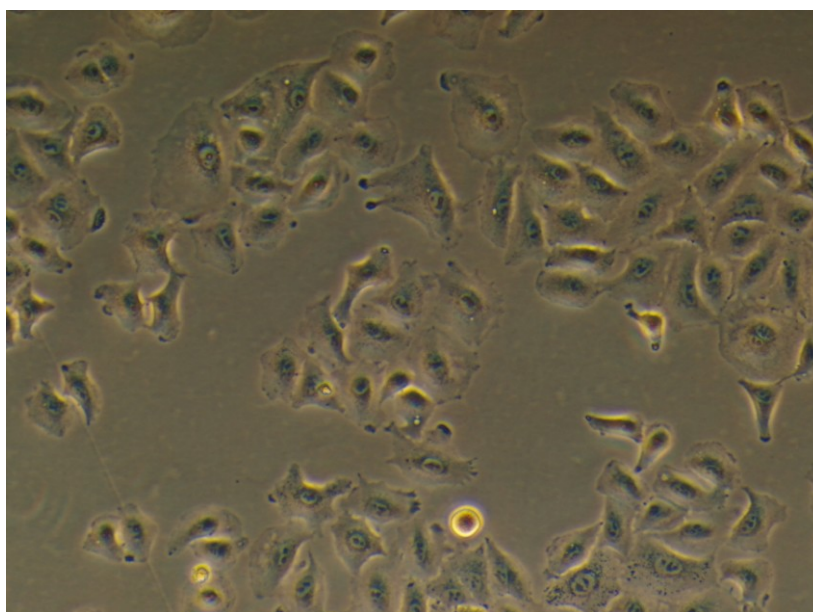
**Figure 3.6** IC<sub>50</sub> calculation showing drug degradation leading to apparently increased Round 0 IC<sub>50</sub> value (12 $\mu$ M) compared to that calculated from multiple experiments with freshly prepared cisplatin (**Figure 3.7**). This is one example of the typical results obtained before realisation of the problems observed with cisplatin (n=2 with SEM error bars). Each cisplatin concentration was performed in triplicate for the dose response and normalisation was done for each plate individually before mean calculation and plotting.



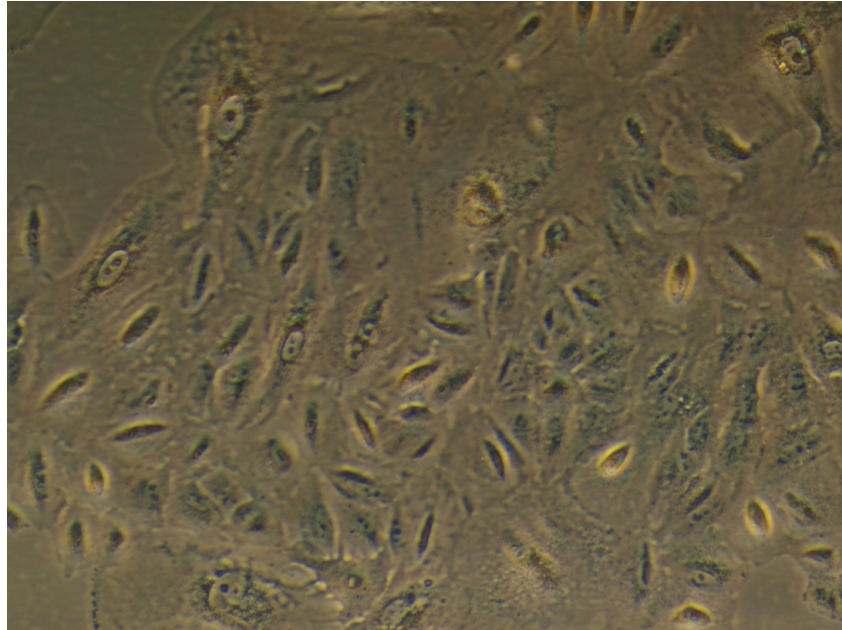
**Figure 3.7** IC50 calculation showing Round 0 IC50 = 5 $\mu$ M (n=6, with SEM error bars). For all these experiments cisplatin was prepared fresh on the day of treatment. Each cisplatin concentration was performed in triplicate for the dose response and normalisation was done for each plate individually before mean calculation and plotting.

### 3.3 Trials of Dosing using Optimised Conditions

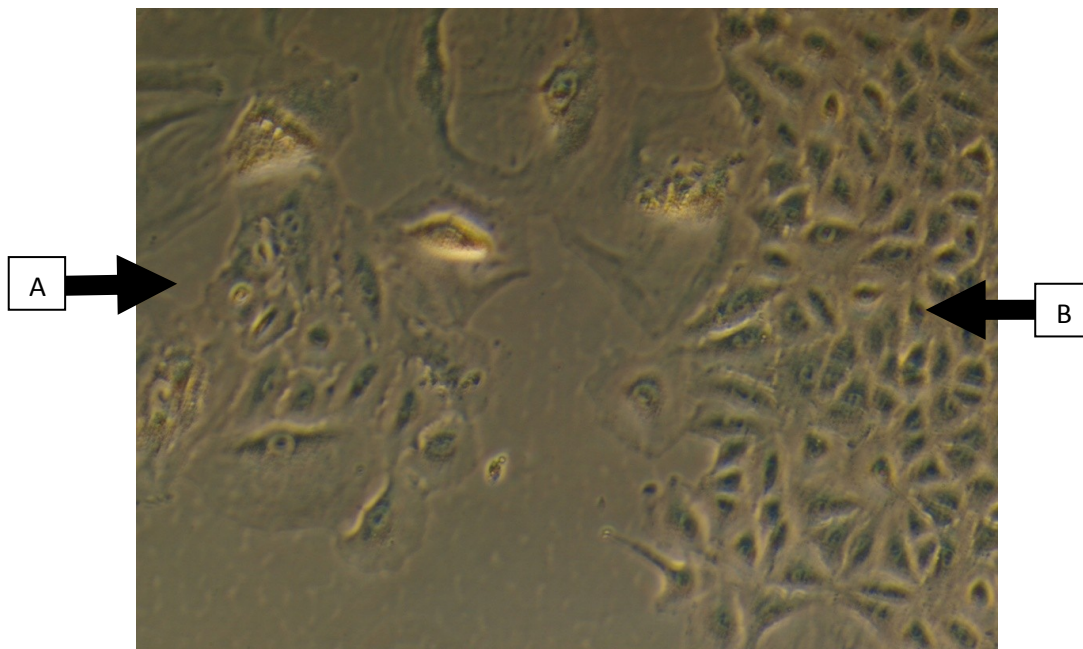
After successful optimisation of the basic protocol for the cisplatin dosing, the rounds of drug dosing of the cells commenced (Chapter 2 Section 2.1.3). Cells were seeded at a density of  $2.0 \times 10^6$  cells per flask (approximately 70% confluency at time of drug treatment) and dosed with the IC<sub>50</sub> value (calculated by the MTT assay) commencing with a concentration of 5  $\mu$ M for the chemo-naïve (Round 0) cells. After 72 hours incubation, the cisplatin was removed and the cells were split, and allowed to recover for various lengths of time: 4 days, 1 week, 2 weeks, 3 weeks and 4 weeks. Recovery was assessed by studying the cells return to normal growth patterns (Figures 3.8 – 3.10). It was observed that normal cellular growth did not restart within one week and therefore, at this stage, four days and 1 week recovery periods were discontinued. Figure 3.8 shows an example of how normal, chemo-naïve cells appear 24 hours after cell splitting. Figure 3.9 shows an example of how stressed cells appear after ten days of recovery whilst Figure 3.10 shows stressed cells with some patches of normal growth occurring, again at ten days recovery.



**Figure 3.8** Normal cells 24 hours after split (seeding density of  $2 \times 10^6$  cells) showing healthy growth, with fairly uniform shape, size and structure. Taken at 10x magnification.

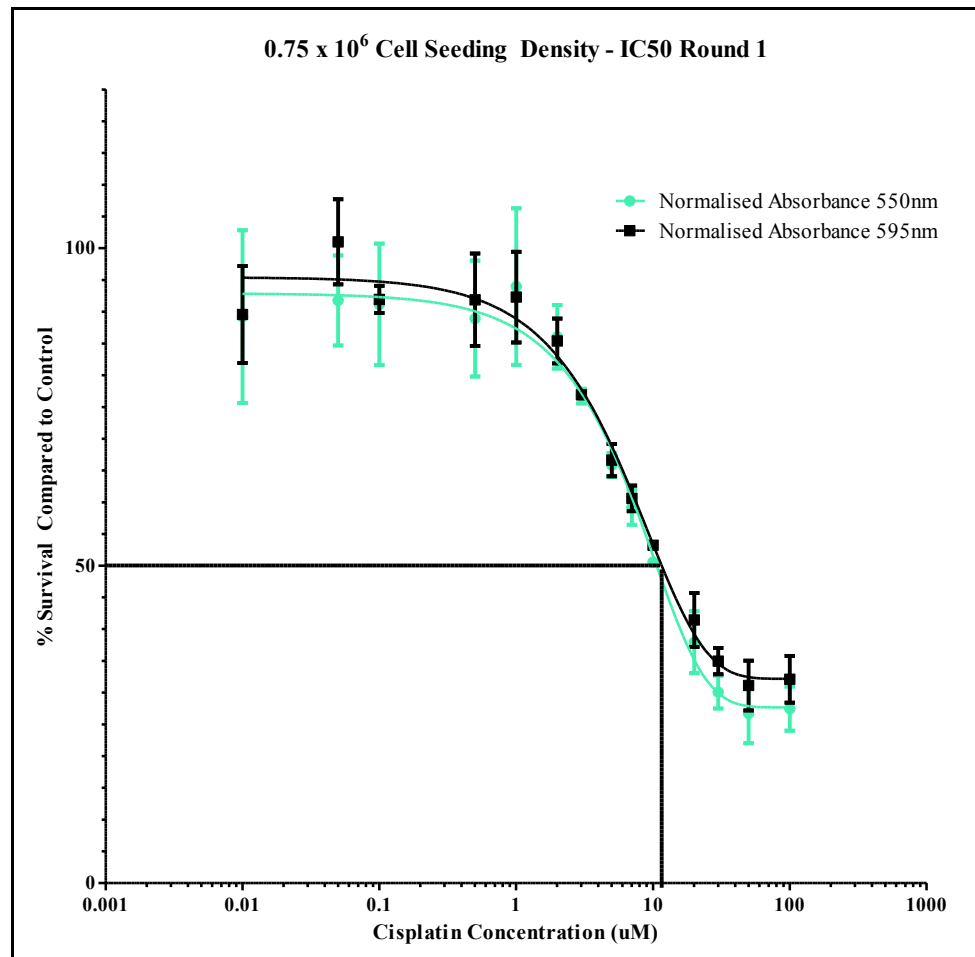


**Figure 3.9** Cells at ten days recovery showing stressed characteristics, including elongation, loss of defined structure and granularisation, which may be indicative of transition from an epithelial to a mesenchymal morphology. Cells were originally seeded at a density of  $2 \times 10^6$ , which was followed by drug dosing with Round 0,  $IC_{50} = 5\mu M$ . Taken at 10x magnification.



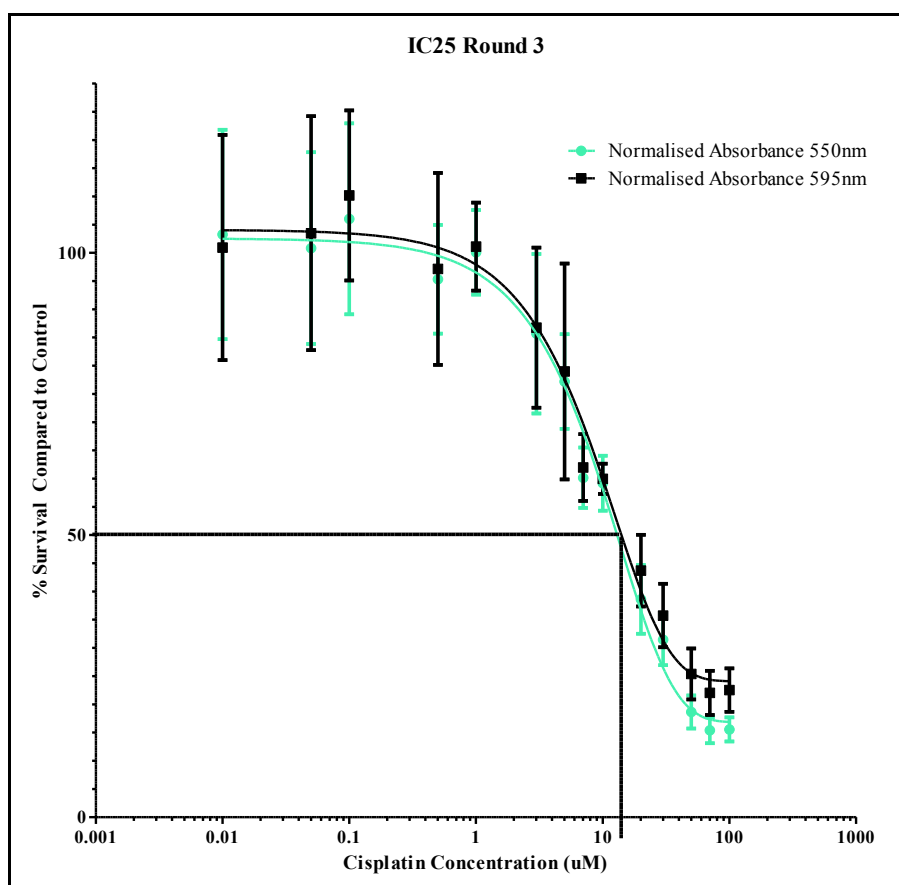
**Figure 3.10** Cells at ten days into recovery. Originally cells were seeded at a density of  $2 \times 10^6$ , which was followed by Cisplatin dosing with the Round 0,  $IC_{50} = 5\mu M$ . Stressed cells can be observed (A) with similar characteristics to those seen in **Figure 3.9**, alongside patches of normal growth (B) similar to that seen in **Figure 3.8**. Taken at 10x magnification.

Cell seeding densities were then tested with the shortest remaining IC50 recovery time (2 weeks) by using  $0.75 \times 10^6$  cells per flask (<50% confluency at drug dosing, Figure 3.11) as well as the  $2 \times 10^6$  cells per flask (70% confluency at drug dosing).



**Figure 3.11**  $0.75 \times 10^6$  seeding density. IC50 Round 1 calculation ( $IC_{50} = 11.5\mu M$ ,  $n=2$ ) after 2 weeks recovery post Round 0  $IC_{50} = 5\mu M$  dosing. Each cisplatin concentration for dose response has been performed in triplicate and normalised to the mean survival of the control cells grown in complete media per plate. Error bars show SEM.

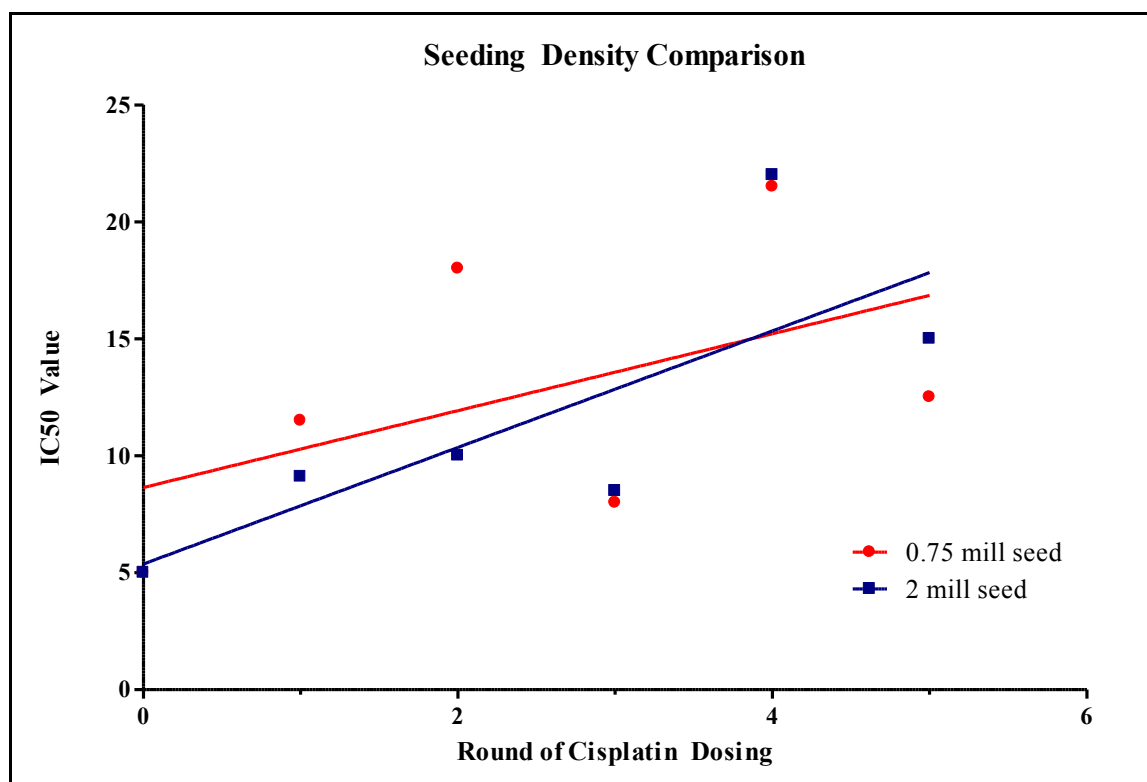
As it had previously been reported in a study of A549 cells that  $IC_{50}$  dosing was too toxic for continued cell growth (Barr and O'Byrne, 2010), and dosing with  $IC_{25}$  concentration was also investigated. Unlike  $IC_{50}$  dosing where 4 days and one week recovery periods did not result in restoration of normal cellular growth, it was found that for  $IC_{25}$  dosing (seeding at  $0.75 \times 10^6$  cells per flask) recovery did occur within 1 week (Figure 3.12).



**Figure 3.12**  $0.75 \times 10^6$  seeding density. IC25 Round 3 calculation (IC25 =  $7\mu\text{M}$ ,  $n=2$ ) after 1 week recovery post Round 2 IC25 =  $4\mu\text{M}$  dosing. Each cisplatin concentration for dose response performed in triplicate and normalised to the mean survival of the control cells grown in complete media per plate. Error bars show SEM.

Comparisons between the two seeding densities tested revealed a gradual (if not significant) increase in IC50 value across the rounds (linear regression analysis in Prism –  $0.75 \times 10^6$  cell seed  $p\text{-value} = 0.3117$ ,  $r^2 = 0.2507$ ;  $2 \times 10^6$  cell seed  $p\text{-value} = 0.0713$ ,  $r^2 = 0.5978$ ). More deviation between the rounds especially for the  $0.75 \times 10^6$  seeding was seen compared to the IC25 values (Figure 3.13).

To investigate any potential differences in resulting IC50 values due to the cell seeding density, a Friedman's test was performed on the data. This showed no significant difference over the period of five completed rounds of cisplatin treatment (Friedman chi-squared = 0.2,  $p\text{-value} = 0.6547$ ).



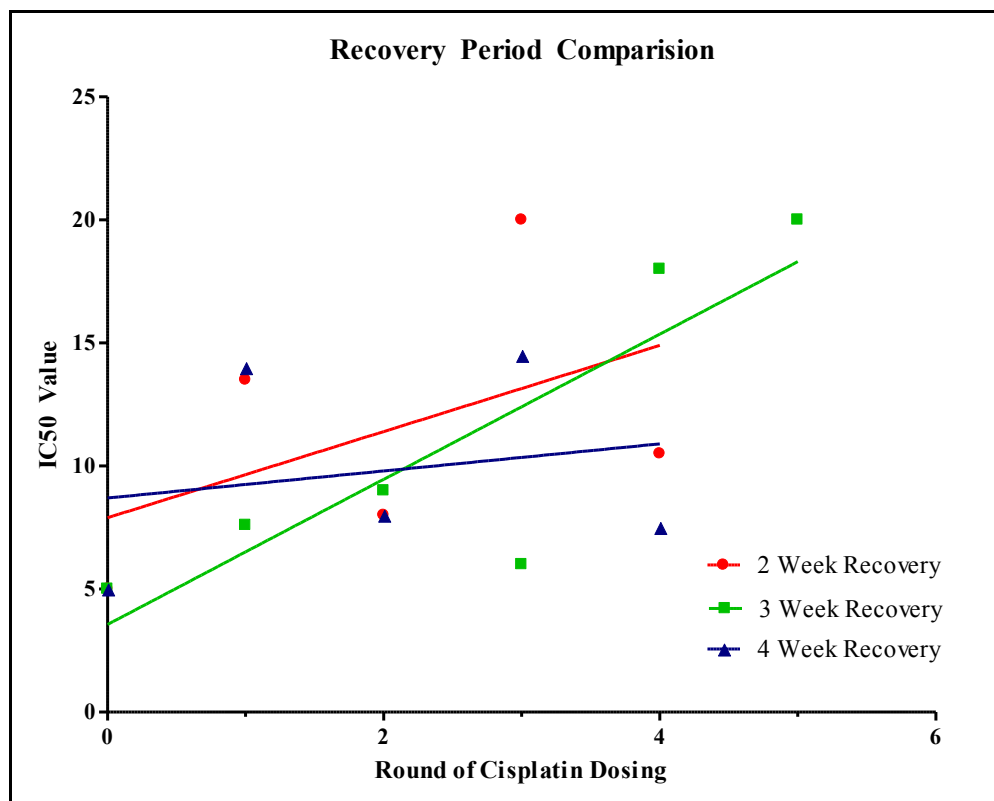
**Figure 3.13** Comparison of  $0.75 \times 10^6$  cell seed and  $2 \times 10^6$  cell seed over five rounds of IC50 dosing. IC50 values plotted from two separate cisplatin dose responses at each round.  $0.75 \times 10^6$  cell seed  $p$ -value = 0.3117,  $r^2 = 0.2507$ ;  $2 \times 10^6$  cell seed  $p$ -value = 0.0713,  $r^2 = 0.5978$ .

When comparing the different recovery periods initially, 2 weeks, 3 weeks and 4 weeks (Figure 3.14), a gradual increase in IC50 was observed for all three. The most notable increase and the only one showing statistical significance was the 3 week recovery, where IC50 increased from  $5\mu\text{M}$  at Round 0 to  $20\mu\text{M}$  at Round 5 ( $p$ -value = 0.0287,  $r^2 = 0.7367$ ). 2 week recovery and 4 week recovery ( $p$ -value = 0.4107,  $r^2 = 0.2325$  and  $p$ -value = 0.7396,  $r^2 = 0.04243$  respectively) had more gradual increases in IC50 values but similar to the results from the seeding density comparison (Figure 3.13), large fluctuations were seen between Rounds 2 and 4.

Using the non-parametric Friedman's test (R Statistics v.15.1), it was found that there was no significant difference between the three different recovery periods (Friedman chi-squared = 0.1333,  $p$ -value = 0.9355) and paired Wilcoxon Signed Rank tests between each of the

regimes (2 week versus 3 week, 3 week versus 4 week, and 2 week versus 4 week) also revealed no statistical significance between the different regimes (2 week versus 3 week –  $p$ -value = 0.8551; 3 week versus 4 week –  $p$ -value = 1; 2 week versus 4 week –  $p$ -value = 0.4227).

Based on these findings, the IC<sub>50</sub> dosing with 3 weeks recovery time course was selected for the model. As it is likely that different dosing schedules may incur different genetic changes over the time of treatment and therefore be useful models of different clinical situations the IC<sub>25</sub> dosing with 1 week recovery was also run in parallel. Both models were run until their IC values had increased four-fold. The mean IC values for all data obtained from the two cell culture models are given in Table 3.2.



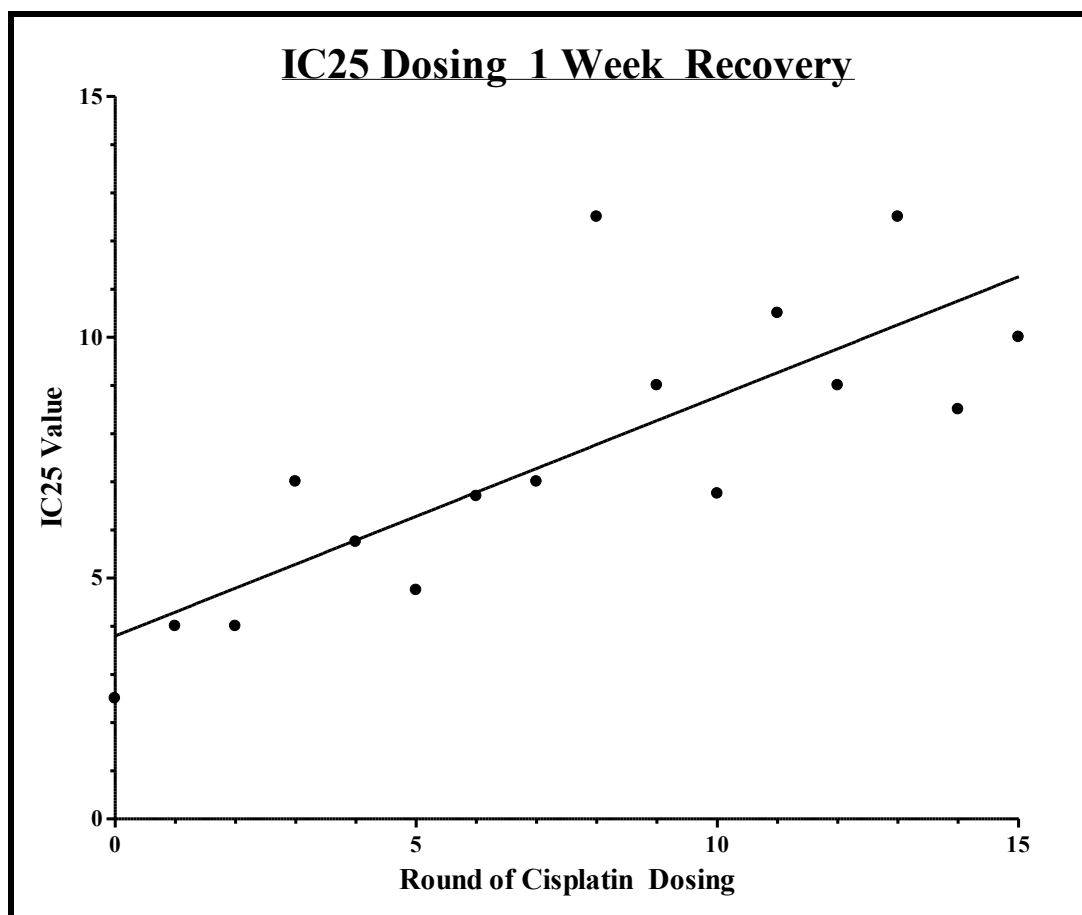
**Figure 3.14** Comparison of 2 week, 3 week and 4 week recovery periods over four to five rounds of IC<sub>50</sub> dosing. IC<sub>50</sub> values plotted from two separate cisplatin dose response curves at each round. Linear regression analysis (Prism); 2 week recovery  $p$ -value = 0.4107,  $r^2$  = 0.2325; 3 week recovery  $p$ -value = 0.0287,  $r^2$  = 0.7367; 4 week recovery  $p$ -value = 0.7396,  $r^2$  = 0.04243.

**Table 3.2** Details of mean IC50 and IC25 values calculated for each round of cisplatin dosing under all conditions tested. All values are given in  $\mu\text{M}$ . See **Figures 3.11** and **3.12** as examples of dose response curves and corresponding IC50 value calculation.

|                                                                   | Round 0 | Round 1 | Round 2 | Round 3 | Round 4 | Round 5 | Round 6 | Round 7 | Round 8 | Round 9 | Round 10 | Round 11 | Round 12 | Round 13 | Round 14 | Round 15 |
|-------------------------------------------------------------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|----------|----------|----------|----------|----------|----------|
| <b>IC25<br/>1 week<br/>recovery</b>                               | 2.5     | 4       | 4       | 7       | 5.75    | 4.75    | 6.7     | 7       | 12.5    | 9       | 6.75     | 10.5     | 9        | 12.5     | 8.5      | 10       |
| <b>0.75 x10<sup>6</sup><br/>cell seed<br/>2 week<br/>recovery</b> | 5       | 11.5    | 18      | 8       | 21.5    | 12.5    |         |         |         |         |          |          |          |          |          |          |
| <b>2 x 10<sup>6</sup> cell<br/>seed<br/>2 week<br/>recovery</b>   | 5       | 9.1     | 10      | 8.5     | 22      | 15      |         |         |         |         |          |          |          |          |          |          |
| <b>2 week<br/>recovery<br/>2 x 10<sup>6</sup> cell<br/>seed</b>   | 5       | 13.5    | 8       | 20      | 10.5    |         |         |         |         |         |          |          |          |          |          |          |
| <b>3 week<br/>recovery<br/>2 x 10<sup>6</sup> cell<br/>seed</b>   | 5       | 7.6     | 9       | 6       | 18      | 14      | 21      | 16      | 18.5    | 30      | 21       |          |          |          |          |          |
| <b>4 week<br/>recovery<br/>2 x 10<sup>6</sup> cell<br/>seed</b>   | 5       | 14      | 8       | 14.5    | 7.5     |         |         |         |         |         |          |          |          |          |          |          |

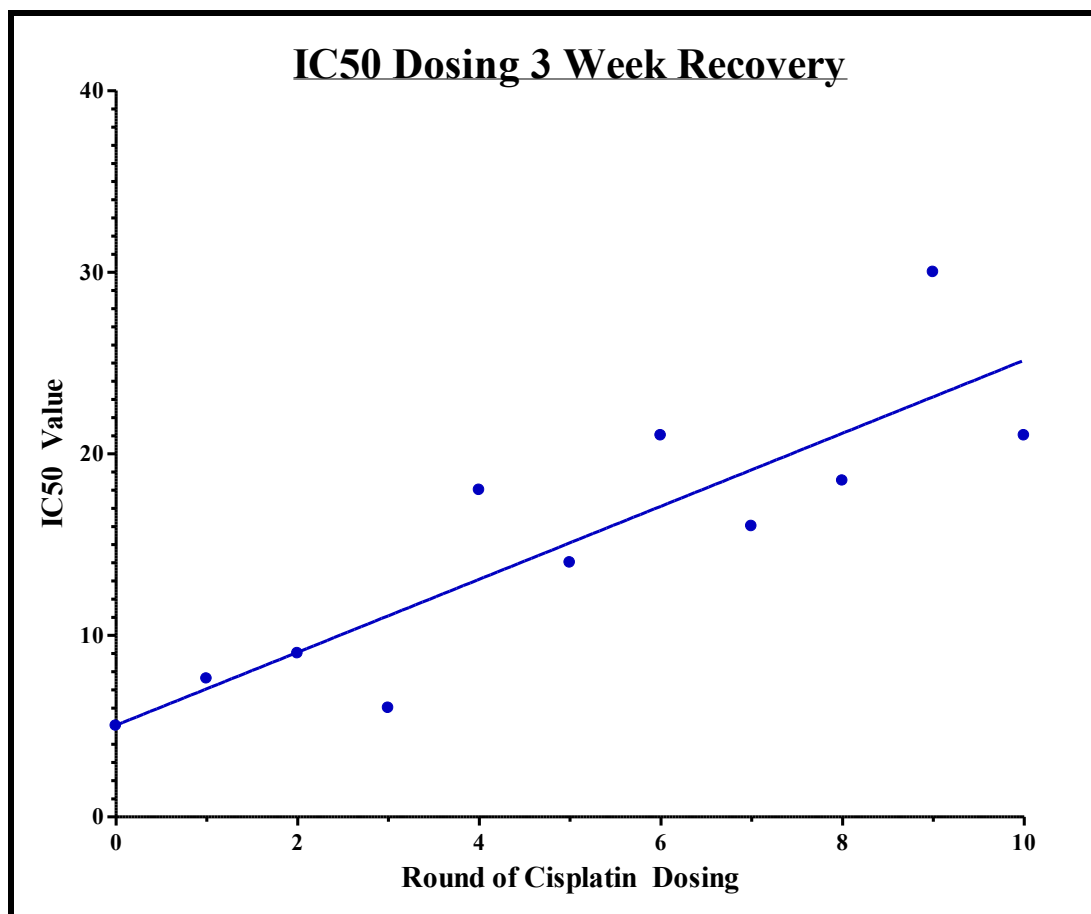
### 3.4 Increasing Cisplatin Resistance

A consistent gradual trend has been observed in the IC25 dose 1 week recovery schedule (Table 3.2 and Figure 3.15) with the increase from IC25 = 2.5 $\mu$ M at Round 0 to IC25 = 10 $\mu$ M after fifteen doses of cisplatin. A statistically significant positive correlation between the number of cisplatin doses and increase in IC25 value was detected (linear regression analysis [Prism, GraphPad],  $p$ -value = 2E-04,  $r^2$  = 0.6342).



**Figure 3.15** Plot of the mean for IC25 values from the cisplatin dose response at each time point from chemo-naïve cells round 0 to round 15. A significant increase in IC25 value can be seen from 2.5 $\mu$ M to 10 $\mu$ M ( $p$ -value = 2E-04,  $r^2$  = 0.63).

A four-fold increase of IC<sub>50</sub> after a total of ten rounds was observed (Figure 3.16) in the IC<sub>50</sub> dose with 3 weeks recovery between each round of cisplatin treatment. Chemo-naïve cells at Round 0 had an IC<sub>50</sub> = 5μM that incremented to 21μM at round 10. As with the IC<sub>25</sub> dose 1 week recovery schedule, a statistically significant positive correlation between the number of cisplatin doses and increase in IC<sub>50</sub> value was detected (linear regression analysis [Prism, GraphPad],  $p$ -value = 6E-04,  $r^2$  = 0.7508).



**Figure 3.16** Plot of the mean for IC<sub>50</sub> values from the cisplatin dose response at each time point from chemo-naïve cells round 0 to round 10. A significant increase in IC<sub>50</sub> value can be seen from 5μM to 21μM ( $p$ -value = 6E-04,  $r^2$  = 0.75).

### 3.5 Chemo-naïve Normal Cell Time Course

Extensive *in vitro* cell culture over a long time period may cause global gene expression changes in its own right and it is important to take these in to consideration. One would anticipate that the largest and most significant fold changes in gene expression in the cell model will be due to the cisplatin exposure due to the selection pressure applied by the presence of the drug during growth over a period of time.

To provide a control for the changing genomics of the A549 cell line over numerous rounds of normal cell culture, a time control was therefore set up. DNA and RNA from chemo-naïve cells were extracted at the same time intervals and approximately the same cell passage as for the cells subjected to successive rounds of cisplatin exposure in the IC50 dose 3 week recovery time course.

### 3.6 Discussion

The development of the two time courses of cisplatin dosing through optimisation of a multitude of variables outlined in this chapter has made it possible to steadily increase the dose of cisplatin required to inhibit cellular growth by 25% or 50%. Many obstacles to the initial set-up of this model and their subsequent solutions will make the products (DNA and RNA extracted at each treatment round) of this robust model of cisplatin treatment useful for future genomic analyses including those presented in the forthcoming chapters of this thesis. Both DNA and RNA from cell aliquots taken from each round of dosing were extracted (as described in Chapter 2 Sections 2.2.3.1 and 2.2.4) at the same time as the next cisplatin dose was applied to the cells. The cells of this next round were then allowed to recover prior to

determining the new IC value. This synchronisation should give an idea of the environment within the cells at the time of treatment.

The final cisplatin time courses chosen involved a 72 hours incubation period with the drug at either IC25 or IC50 concentration. After this incubation the cisplatin was removed and the cells allowed to recover before the next IC25 or IC50 concentration was calculated, either 1 week or 3 weeks later, respectively. Subsequent to the establishment of this cellular model, a paper has been published that suggests that 24 hours after 24 hours of cisplatin treatment would be a suitable point at which to investigate the effects of cisplatin within in the cell (Barr *et al.*, 2013). In this paper they explored the biochemical and physiological properties of cisplatin within resistant cell lines compared to age-matched parent cell lines and discovered some interesting findings. In comparisons between resistant cell lines compared with the equivalent parent, chemo-naïve cells, those exposed to cisplatin constantly overcame the drug's effects through reduced cisplatin-GpG adduct formation, enhanced DNA double strand breakage repair, decreased cisplatin induced apoptosis, and reduced cellular uptake of cisplatin. There also appeared to be an accumulation of cells in the G0/G1 phase of the cell cycle in the cisplatin-resistant cell lines compared to the parents, showing a move towards a stem cell phenotype with markers of EMT up-regulated. This will be discussed further in Chapter 7 Section 7.1. It would be interesting to compare the differences, if any, between varied incubation times with the drug. A possible expansion of this model, combining the methods from the Barr *et al.* paper (2013) would be to perform shorter incubations with cisplatin, between 6 and 24 hours, and to extract DNA and RNA immediately after the incubation was complete. The immediate surviving cells from these short incubations would hopefully contain more viable cells and so be amenable to DNA and RNA extraction at that time.

There will be many points to consider when exploring the data from these treated cells as evolutionary populations have been generated over the time courses. When cells are split, be it for dose response curves, cisplatin treatment, DNA or RNA extraction, or flask seeding for the continuation of the resistant line, the cell will not be genetically uniform. The underlying differences between the cells, though probably small, may cause some of the variation, seen in IC50 and IC25 values, observed at each time point (Figures 3.15 and 3.16). In some ways, this slight genetic diversity in the model may mimic the genetic heterogeneity of a tumour in the clinical setting as not every cell will have exactly the same genome. It is also possible that certain cells with driver phenotypes may proliferate and recover more quickly than others and thereby constitute a larger proportion of the cells in a given flask or well. This again may be representative of a given tumour phenotype as clonal selection may occur in the cellular population.

The heterogeneity introduced in this model could have been avoided by direct, deliberate clonal selection at each cell split. This may have been useful as it has been documented that, even with a small number of cells, average measurements of the population may not be representative of any individual cell (Wang and Bodovitz, 2010). Some success has recently been had in the field of single cell genomics (Zong *et al.*, 2012) but these techniques are in their early stages and single cell selection, whole genome amplification, and elimination of contamination are experimentally challenging. As these methods become more main stream the importance of this field of research cannot be overlooked as the heterogeneity within a single tumour could be a key aspect to clinical management and treatment options, as well as potential for diagnostic usage, such as circulating tumour cells in the blood. Even as this technology becomes more widely available and more in depth, it would still generate a

demanding amount of work before the dose response, DNA and RNA extraction, and flask seeding procedures could be completed.

Historically selection of cells or colonies has been done using immunological markers for stable cell line production, such as for transfected cells (Lu *et al.*, 2011). This could have been done here to ensure a single surviving colony was used in the subsequent rounds of cisplatin dosing, DNA and RNA extraction, and dose response calculations. The selection of a single immunological phenotype would however not necessarily be selective enough to guarantee that the resulting cells were genetically identical. In order to produce perfect genetic clones, as the aims of this work would prefer, more complex and technically difficult methods would be required.

This cellular model has its limitations and some of the general ones have already been discussed earlier (see Section 3.1). The mutation profile of untreated A549 cells (Table 3.1) could bias any further mutations acquired during the course of cisplatin treatment. It could be postulated that the mutations of potential drivers, *STK11* and *KRAS*, may encourage other cooperative genes from within the same signalling pathway to be preferentially altered. Considering this, it would be beneficial to expand the dataset further by repeating the optimisation and dosing schedule with other lung cancer cell lines that have different, ideally known clinically common baseline mutations prior to treatment.

Patients being treated with cisplatin in a clinical setting generally undergo six cycles each consisting of a single dose, between 50 to 100 mg/m<sup>2</sup> intravenously over 6 to 8 hours, with three to four weeks recovery between cycles (Chapter 1 Section 1.5 and <http://www.medicines.org.uk/emc/medicine/623/SPC/Cisplatin+1+mg+ml+Sterile+Concentrate>).

In this model of cisplatin resistance, the concentrations of cisplatin the cells were exposed to initially at round 0 (IC<sub>50</sub> value = 5µM, IC<sub>25</sub> value = 2.5µM) equate to very similar concentrations in the area of the flasks used (0.75-1.5 mg in a 150cm<sup>2</sup> flask), which is equivalent to the same 50 mg/m<sup>2</sup> for the IC<sub>25</sub> first cisplatin dosing and 100 mg/m<sup>2</sup> for the first IC<sub>50</sub> dosing. The use of 3 weeks recovery for the IC<sub>50</sub> dosing time course also reflects well the recovery time between chemotherapy cycles in the clinical setting. The two key differences are the deliberate increase in cisplatin concentration over time, done to try and drive resistance relatively quickly, and the lengthened exposure to cisplatin, 72 hours versus 6-8 hours in a clinical setting, the latter again being done to ensure selection for resistance.

An additional consideration is that cisplatin is generally combined with another anti-mitotic (see Chapter 1 Section 1.5) when used to treat patients in the clinic. Different cisplatin combination therapies may lead to different outcomes for different subtypes of NSCLC, such as pemetrexed combined with cisplatin that appears to be less active in lung squamous cell carcinomas than in other types of disease, leading to longer progression-free survival and overall survival (Scagliotti *et al.*, 2009a). Modification of this established A549 cell model with differing drug combinations is another potential option for future investigations.

Despite these recognised potential limitations to this model, cisplatin-resistance has been induced into this A549 cell line over a period of time and, with the procurement of DNA and RNA at each time point, has provided a wealth of data for genomic investigations over the two treated time courses. It allows for investigations over the entire period of cisplatin treatment, not just simple comparisons between parent/chemo-naïve versus resistant phenotypes. This model is a first step in dissecting the effects of cisplatin over the course of a treatment regime and may help in the understanding of the genetic and genomic mechanisms underpinning cisplatin resistance.

## **4 Global Gene Expression Analysis of Cisplatin Resistance**

### **4.1 Introduction to Global Gene Expression Analysis**

Global gene expression profiling allows examination of genome activity across the whole transcriptome. Understanding of the diversity of transcripts in and between individuals is evolving and becoming more complex, with transcript variants now known to include canonical, alternatively spliced, alternative start sites, alternative polyadenylation, retained introns, repeated exons, antisense transcription, chimeric transcripts and non-coding RNAs. All of these RNA variants may alter the phenotype of the sample of interest by causing changes at the protein level.

It seems inevitable that an external stimulus, such as the dosing of cells with rounds of cisplatin, would have pronounced effects at the mRNA level leading to a variety of changes in the cells' phenotype.

There are various methods of exploring gene expression, such as quantitative reverse transcriptase polymerase chain reaction (qRT-PCR), expression microarrays and RNA sequencing (RNA-seq), each with their own benefits and pitfalls. In this work, global gene expression analysis was performed using microarray technology to gain a relatively comprehensive overview of the transcriptome (Fryer *et al.*, 2002).

Quantitative Real Time PCR (qRT-PCR) is useful to study specific genes and their behaviour under certain circumstances. It amplifies and quantifies the target cDNA molecule simultaneously, which is generated from the RNA isolated from tissues or cells of interest. It is a simple, accurate and potentially fully quantitative technique that can be used to focus on a selection of genes, for example, within a pathway or network of interest, although it is

essential to normalise the data to genes that do not alter their expression under different conditions. Consequently, the approach is not hypothesis-free – genes are selected as candidates based on their potential functions – and only a limited number of genes can be studied due to cost implications and the quantity of RNA required for the procedure, if the entire genome were to be investigated.

RNA-seq is the new frontier in gene expression profiling. With the explosion in availability of Next Generation Sequencing, it has become feasible to sequence all transcripts within a sample. Theoretically it provides limitless sensitivity, accuracy and a broad dynamic range. Unlike qRT-PCR, RNA-seq is a hypothesis-neutral approach that has the capability to discover novel transcripts and isoforms as well as determining gene structure and sequence in one experimental protocol. The technology at present however is expensive, both in cost and RNA quantity required, and some studies have observed a GC bias in RNA-seq data. There may also be mapping ambiguity for paralogous sequences and an erroneous association between gene length and differential expression (Oshlack *et al.*, 2010).

Gene microarrays have been used here to study the global transcriptome. Historically, microarrays have been on the market longer than RNA-seq and so have improved drastically in content and robustness, as well as significant reduction in their cost. They allow comparisons between samples over tens of thousands of genes. Each sample is labelled with a different fluorescent dye and hybridised to the probes on the array. After multiple washes, the hybridised slide is scanned and the relative fluorescence of each probe is measured.

In similarity with RNA-seq, microarrays are another hypothesis-neutral way of exploring the transcriptome. They also allow high-throughput processing of multiple samples, by automated hybridisation and washing steps for several arrays simultaneously. Further development and miniaturisation has also allowed for automated scanning through the use of

array plates. Microarrays are a relatively cheap, fast method of measuring expression levels of large numbers of genes simultaneously, however only known transcripts are detectable and low expression transcripts may not be identified.

## **4.2 RNA Preparation: Extraction and Quality Control**

### **4.2.1 RNA Extraction from Cultured Cells**

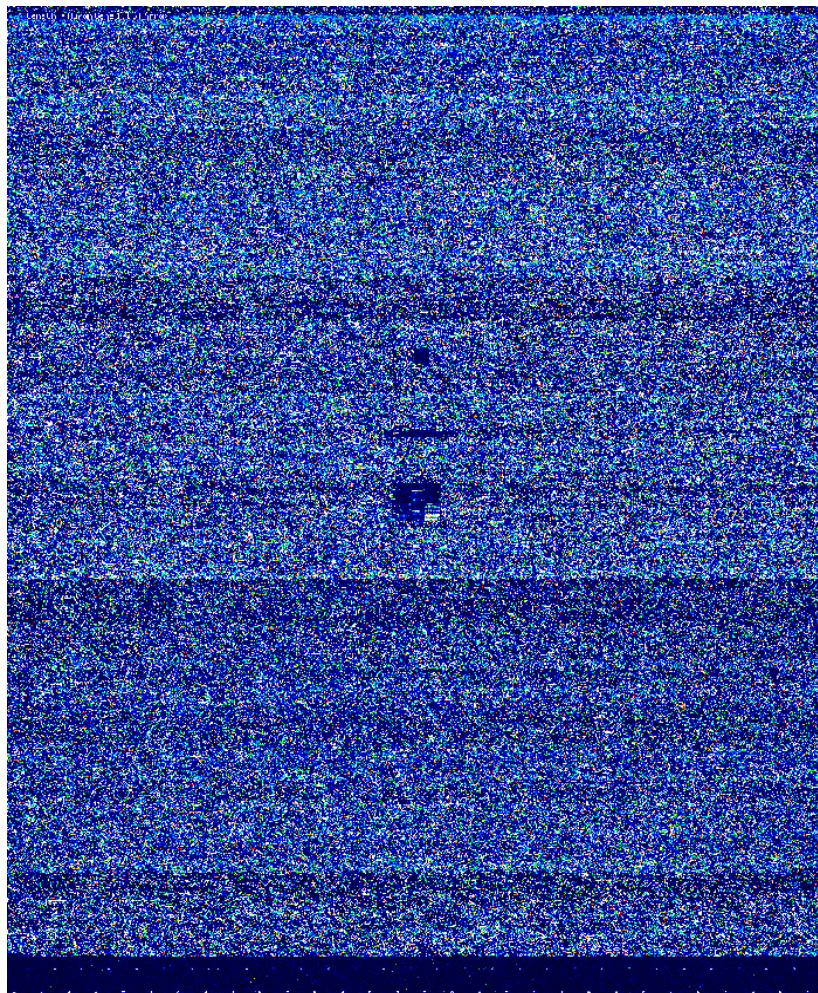
RNA was extracted from homogenised, lysed A549 cells from each cisplatin dosing round and quantified and quality control checked as outlined in Chapter 2 Section 2.2.3.1. Every effort was made to ensure RNases did not interfere with extraction however on quantification of the triplicate samples for round 2 of the IC25 Dose 1 week recovery time course, negligible RNA was present. Consequently RNAs from this round were not included in the downstream analyses. All RNAs from cultured cells had RINs  $\geq 9.5$ .

### **4.2.2 RNA Preparation for Microarray Hybridisation**

Extracted and quantified RNA samples were prepared for and subject to microarray hybridisation, washing and scanning using the GeneTitan®, as outlined in Chapter 2 Section 2.2.3.2. Any voids caused by bubbles or other artefacts would lead to areas on the array with no probe-sample hybridisation and so particular care was taken when setting up the hybridisation tray.

During the scanning process, the scanner uses markers on the corners of each array to ensure all are aligned before their fluorescence is measured. This produces a CEL file, which stores

the results of intensity calculations that originate from the fluorescence reading. Ideally, the intensity of the fluorescence should be even across the whole of the array with no patches of overly bright or dim areas indicating even hybridisation, and no bubbles, voids or other artefacts. The fluorescence measurements are read by the scanner from a greyscale image; the false-colouring may make it easier to identify areas of sub-optimal hybridisation (Figure 4.1). Combining this CEL file with the manufacturer's information about specific probes and their locations on each array, fluorescence can be assigned to each probe and hence its target transcript. All samples passed this first step of quality control and were therefore taken forward into the expression analysis.



**Figure 4.1** An example of a false-coloured hybridised array as scanned using the GeneTitan. There is no bias towards a single colour and intensity is relatively uniform across its entirety, with no extremely bright or dim areas on the array implying even hybridisation.

#### 4.2.2 Microarray Quality Control, Normalisation and Transformation

Quality control is vital in microarray experiments as it allows outliers to be detected, excluded and repeated if required. A single array on a plate may appear vastly different from all the others on that plate; they may appear as an obvious outlier on a log expression signal box and whisker plot (Sections 4.2.2.1 and 4.2.2.2).

Normalisation between the arrays is required to correct for systematic differences between the samples and the arrays themselves that are not due to biological differences, such as overall spot intensity, array location on a plate, origin of a plate or scanning parameters. Statistical batch correction may be required hence the need for careful batch recording through the experimental processes (Chapter 2 Section 2.2.3.2). It also makes distributions comparable between all samples of interest. Robust Multiarray Averaging (RMA) normalisation was used for all datasets. Log transformation was also implemented to remove skew within the distributions and makes relative changes less sensible to errors.

The CEL files (raw data) generated by the GeneTitan from the Affymetrix gene arrays were subject to probeset summarisation by using Affymetrix Power Tools (APT). These are a set of cross-platform command line programs that implement algorithms for analysing and working with Affymetrix microarray data. Multiple summarisation methods are available through APT, including Probe Logarithmic Intensity Error (PLIER) and RMA. Using APT allows one to:-

- Conduct quantile normalisation using a subset (sketch) of the data which results in smaller usage.

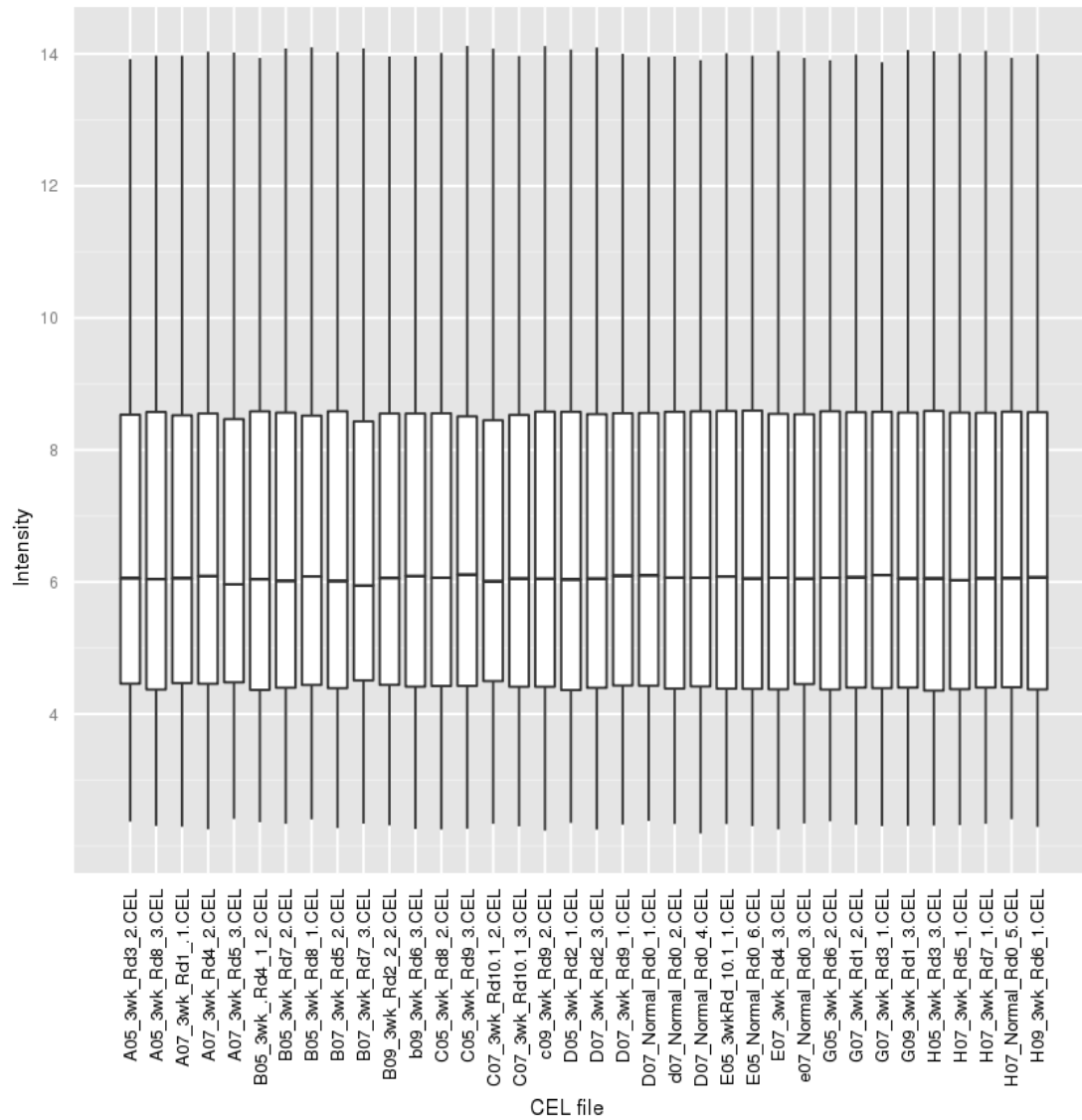
- Analyse microarray data in probesets, which when combined with the sketch quantile normalisation allows the analysis of thousands of chips on a single computer or in parallel across a cluster of computers.
- Save target normalisations and probe (feature) effects for later use.
- Have the ability to group probesets into larger “meta” probesets as specified by the user via text file.

APT also enables combination with databases, for example Affymetrix array probe annotation files, that retain information about the probes on each array.

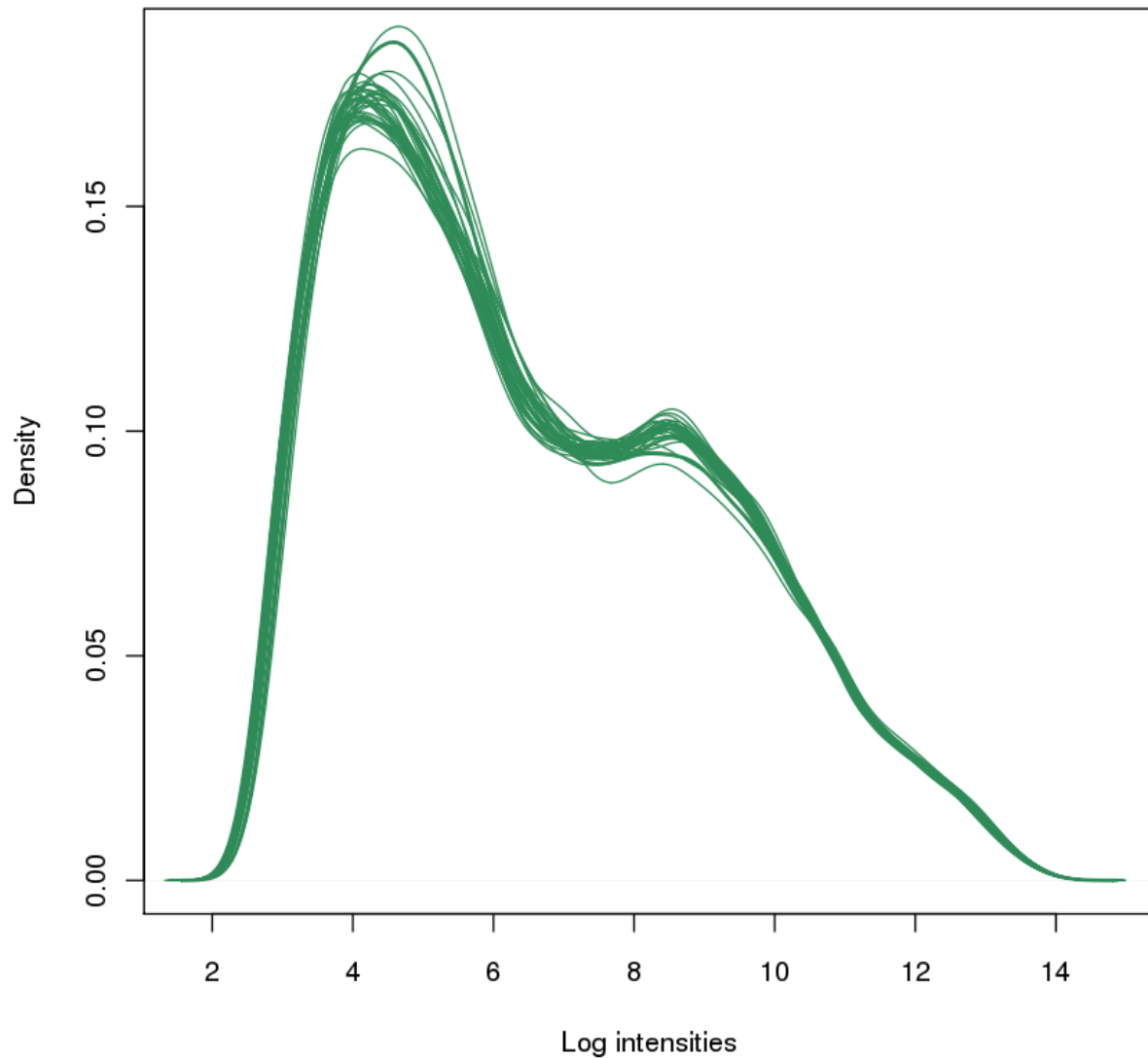
After processing all the microarrays through the APT algorithm, the resulting data set was subjected to further quality control measures. R and its associated packages “Affy” and “Limma” were used to generate box plots of expression intensities, hierarchical clustering plots and signal density distributions. Affymetrix software Expression Console was also used to calculate a variety of quality assessment metrics. After this additional round of quality control, differential expression analyses were conducted as detailed in Chapter 2 Section 2.2.3.4.

#### *4.2.2.1 IC50 Dosing 3 Week Recovery*

Quality control measures, conducted in R Statistics v. 2.15.0 and its associated packages, revealed that no outliers existed for the IC50 Dosing 3 week recovery time course as illustrated by the box whisker plots and signal distribution histograms (Figures 4.2 and 4.3).

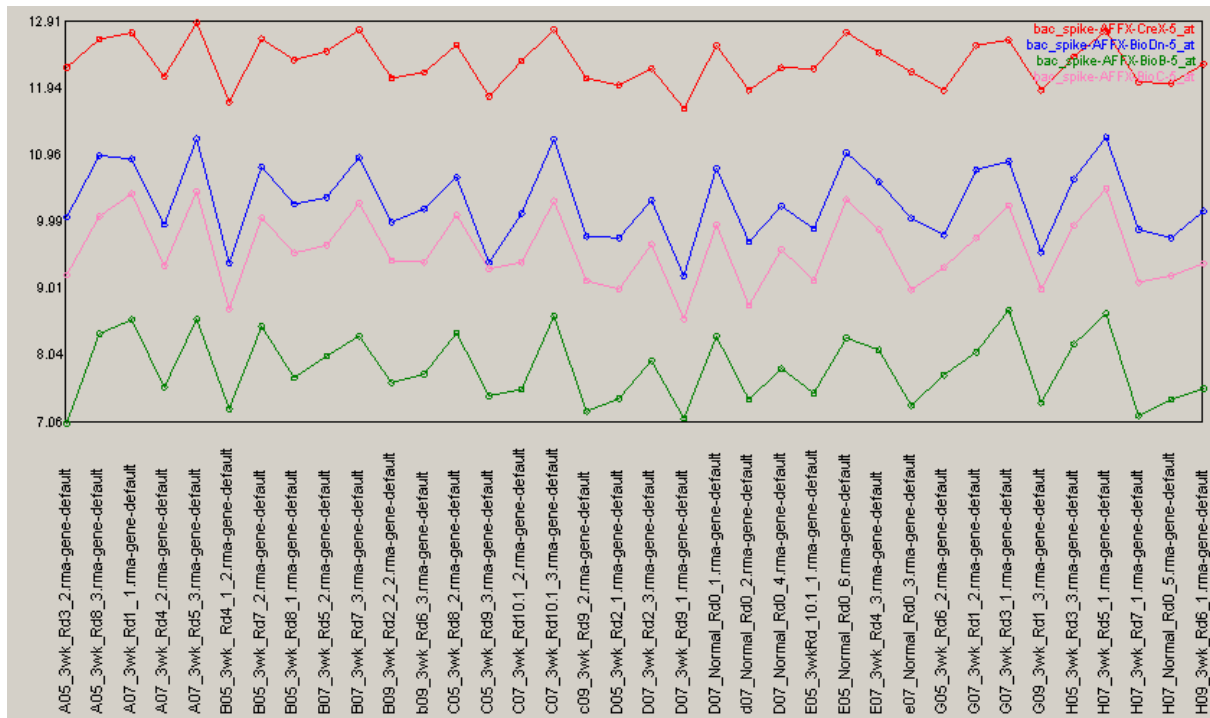


**Figure 4.2** Log Expression Signal showing the box and whisker plots for each array file in the IC50 dose 3 week recovery time course. It can be seen that no outliers exist as all 36 arrays have approximately the same median value with similar interquartile and total ranges.



**Figure 4.3** Signal Distribution Histogram for all 36 arrays from the IC50 dose 3 week recovery time course. The pattern of the distributions for all the arrays is almost identical with no marked outliers.

Individual probe set signal values, including the bacterial spike and polyA spoke probe sets were used as controls. Bacterial spike is the set of probe sets which hybridise to the pre-labelled bacterial spike controls (BioB, BioC, BioD, and CreX) and it is used in identifying problems with the hybridisation and/or array. Each is spiked with a different concentration of fluorescence (BioB the lowest to CreX the highest). If they do not appear in the correct order when comparing intensities i.e. CreX the highest values and BioB the lowest values, it is an indication of hybridisation problems (example shown in Figure 4.4).



**Figure 4.4** Bacterial spike signal values generated in Expression Console, showing each array from the IC50 dose 3 week recovery time course and its intensity for each bacterial spike hybridisation control. It can be seen that, as expected, the CreX control (red) has the highest intensity, followed by BioD (blue), BioC (pink) and BioB (green) respectively.

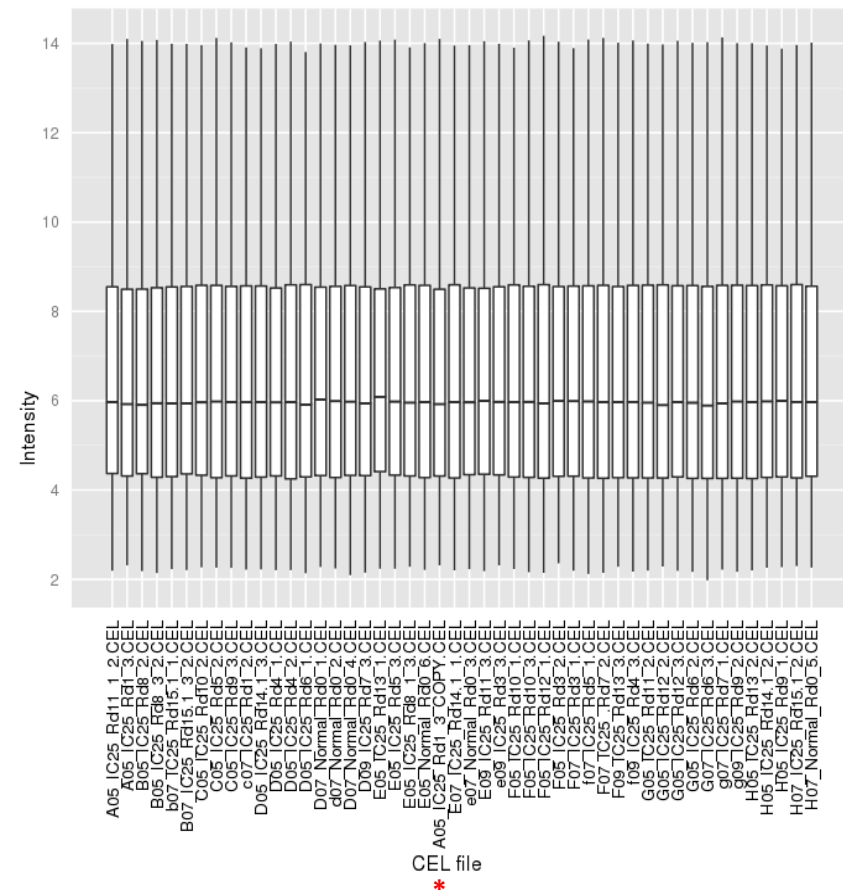
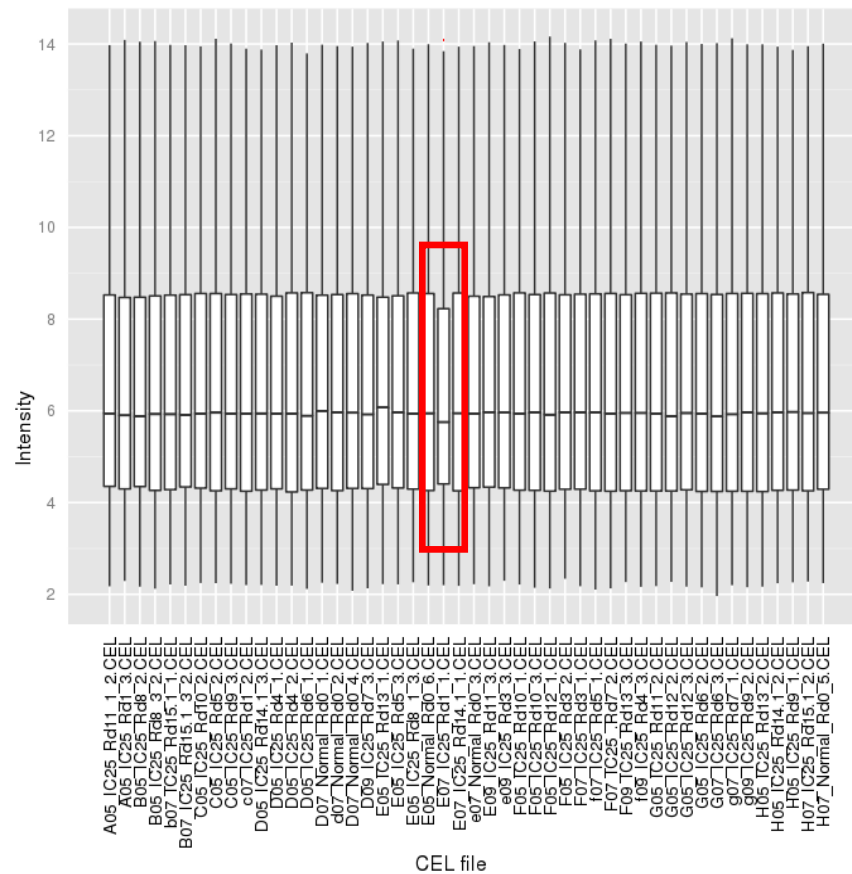
The PolyA spike is the set of polyadenylated RNA spikes (Lys, Phe, Thr and Dap) and it can be helpful in identifying problems with target RNA preparation. Again, each of these four controls are spiked with difference concentrations of fluorescence and on scanning should appear in order of intensity; from Dap with the highest expected intensity, followed by Thr, Phe, and Lys respectively.

#### *4.2.2.2 IC25 Dosing 1 Week Recovery*

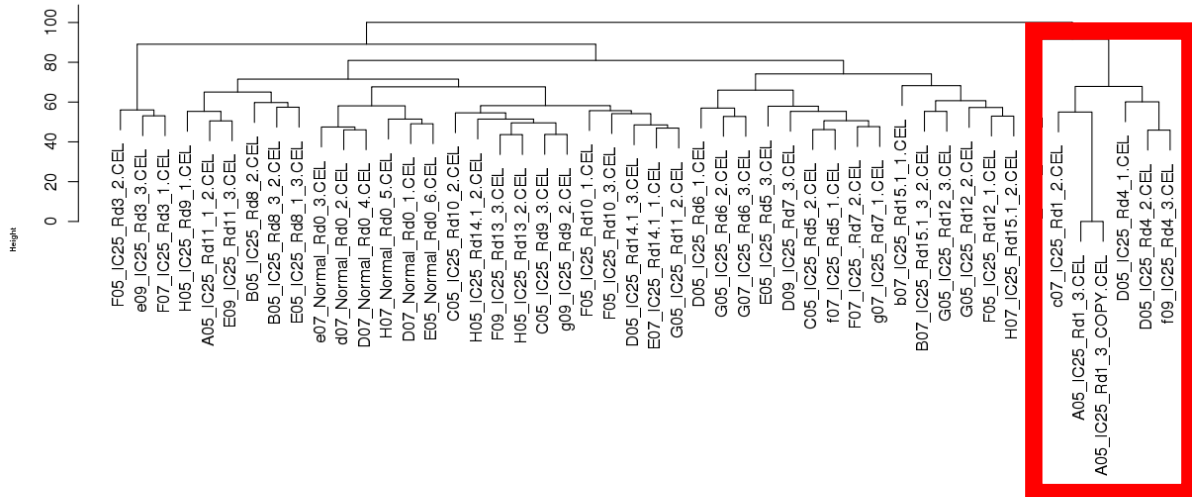
Quality control analyses were performed as outlined in Section 4.2.2.1. These highlighted a potential problem with replicate 1 from Round 1; the box and whisker of log expression signal of the arrays suggest that it may be an outlier. To overcome this problem, a duplicate of the CEL file for replicate 3 was used and the quality control analyses rerun (Figure 4.5).

In order for the two time courses to be comparable and simplify the analyses, a further four rounds were chosen for exclusion so the 15 rounds from the IC25 dose 1 week recovery were reduced to 11 (including round 0) in order to match the IC50 dose 3 week recovery time course. The increased number of comparisons that would result from contrasting all the 15 rounds in the IC25 dose 1 week recovery time course could also lead to a greater chance of potentially falsely statistically significant results, simply due to conducting more comparisons.

Rounds 1, 4, 8 and 12 were chosen for exclusion to take into consideration the prior hierarchical clustering (Figure 4.6) in which rounds 1 and 4 cluster together, slightly removed from the other rounds, and to give a well-balanced spread of observations across the entire time course (rounds 8 and 12). This pruning of the IC25 dose 1 week recovery time course made no significant changes to the conclusions of the experiments.



**Figure 4.5 Left:** Log Expression Signal of IC25 dose 1 week recovery. It can be seen that one box plot does not conform with the others as it has a smaller box and a lower median value, highlighted by red box. **Right:** Log Expression Signal of IC25 dose 1 week recovery. On replacement of the Round 1 Replicate 1 outlier array with a copy of another replicate from the same cisplatin round (highlighted with a red asterisk), the quality of the data taken forward is improved as all box and whisker plots have a similar median, interquartile range and total range.



Hierarchical cluster analysis on a distance matrix computed using an euclidean distance measure

**Figure 4.6** Hierarchical clustering plot of expression data for all time points for the IC25 dose 1 week recovery time course. It can be seen that rounds 1 and 4 appear as outliers.

### 4.3 Differential Gene Expression Analysis

Differential expression analysis is vital to identify significant changes in gene expression, both up- and down-regulated, across the whole time course. In the case of a time course drug dosing, it is worth investigating the dynamic process of genetic response within the cells as those genes that change expression in line with the cytotoxic drug may be causing changes in the cell to allow it to survive. These analyses will lead to further investigations such as network analysis that summarise the co-ordinated activity of the transcriptome (Chapter 5 Section 5.2).

For the differential expression over time analysis, each time point was contrasted with every other within each time course. In order to examine all potential changes in both the IC50 dose 3 week recovery and the IC25 dose 1 week recovery time courses, a moderated F-statistic

was created for each gene across the entirety of both time courses (Chapter 2 Section 2.2.3.4), without simply focusing on start versus resistance end point alone.

Using a significance threshold of  $1E-08$ , the F-statistic and Benjamini-Hochberg (B-H) adjusted  $p$ -values reduced the 33,297 transcript clusters interrogated by the microarray to identify 1,515 transcripts that were significantly altered during the IC50 dose 3 week recovery time course and 1,474 transcripts significantly altered during the IC25 dose 1 week recovery time course.

#### 4.3.1 Top Differentially Expressed Genes from the Two Treated Time Courses

Of the 1,515 significantly differentially expressed genes from the IC50 dose 3 week recovery time course, 183 genes were found to have a greater than three-fold increase or decrease in expression when comparing round 0 with the final round 10. Fifty nine of the genes had a greater than five-fold change in expression. It is interesting that the majority of these 183 genes with large fold changes appear to decrease over the time course with approximately a third (67) showing an increase in expression. Overall, 673 genes were found to have increased fold change expression on comparison between rounds 0 and 10, while 842 had decreased fold change over the same period.

Table 4.1 shows the top 25 differentially expressed genes (for a more comprehensive list [288 gene with B-H adjusted  $p$ -value less than or equal to  $1E-14$ ], see E-Appendix 9.1.1). The top ten differentially expressed genes perform a variety of functions; a complement protein and a cytokine (C3 and GREM1 respectively), metabolism enzymes involved in the urea cycle (CPS1) and in the liver (BAAT), a heparin binding protein (C21orf63), an adhesion molecule (LUM), a metalloprotease (TLL1), a small nuclear RNA (snRNA, ENST00000517058

located on chromosome 5), a folate pathway gene (CRISPLD1), and an inhibitor of fibrinolysis (SERPINE1) (<http://www.ncbi.nlm.nih.gov/gene>).

Although these top ten genes do not appear to be related to the lungs or cancer, other genes (for full listing see E-Appendix 9.1.1) do have more obvious links. Solute carrier family 6 (amino acid transporter) member 14 (*SLC6A14*) is a specific amino acid transporter that has been identified as a novel drug target for oestrogen receptor-positive breast cancers (Karunakaran *et al.*, 2011). Mal, T-cell differentiation protein 2 (MAL2) is a transporter protein that has also been linked with breast cancer (Shehata *et al.*, 2008) and shows an approximate 20-fold decrease between the start and end of the time course. Angiogenin (*ANG*) levels have been indirectly associated with tumour size through a correlation with increasing vascular endothelial growth factor (*VEGF*) concentration (Rozman *et al.*, 2012).

**Table 4.1** IC50 dose 3 week recovery; top 25 differentially expressed genes. Fold change (increase in green, decrease in red) shows change in expression when comparing the chemo-naïve (round 0) cells with those from the end of the time course (round 10).

| ID      | Fold Change Rd0v10 | F Statistic | B-H adj. <i>p</i> .Value | Gene Symbol                               | > 3fold? | > 5fold? |
|---------|--------------------|-------------|--------------------------|-------------------------------------------|----------|----------|
| 8048026 | 1.25               | 518.14      | 4.54E-25                 | <i>CPS1</i>                               |          |          |
| 8033257 | 0.10               | 429.02      | 3.08E-24                 | <i>C3</i>                                 | Yes      | Yes      |
| 8146967 | 0.03               | 417.97      | 3.08E-24                 | <i>CRISPLD1</i>                           | Yes      | Yes      |
| 8162870 | 13.09              | 410.80      | 3.08E-24                 | <i>BAAT</i>                               | Yes      | Yes      |
| 8068202 | 0.13               | 398.05      | 3.85E-24                 | <i>C21orf63</i>                           | Yes      | Yes      |
| 8104654 | 0.04               | 392.98      | 3.85E-24                 | <i>ENST00000517058</i> ,<br>snRNA on Chr5 | Yes      | Yes      |
| 7965403 | 20.11              | 369.96      | 7.77E-24                 | <i>LUM</i>                                | Yes      | Yes      |
| 8098214 | 0.08               | 312.96      | 7.29E-23                 | <i>TLL1</i>                               | Yes      | Yes      |
| 8135069 | 1.90               | 305.44      | 9.14E-23                 | <i>SERPINE1</i>                           |          |          |
| 7982377 | 22.35              | 292.05      | 1.55E-22                 | <i>GREM1</i>                              | Yes      | Yes      |
| 8029086 | 0.07               | 280.06      | 2.35E-22                 | <i>CEACAM5</i>                            | Yes      | Yes      |
| 8148040 | 0.05               | 279.97      | 2.35E-22                 | <i>MAL2</i>                               | Yes      | Yes      |
| 7973084 | 0.97               | 246.00      | 1.35E-21                 | <i>ANG</i>                                |          |          |
| 8114920 | 0.61               | 242.75      | 1.42E-21                 | <i>DPYSL3</i>                             |          |          |
| 8169504 | 0.02               | 242.63      | 1.42E-21                 | <i>SLC6A14</i>                            | Yes      | Yes      |
| 8125530 | 2.59               | 233.96      | 2.23E-21                 | <i>HLA-DMB</i>                            |          |          |
| 8142585 | 0.20               | 223.22      | 4.07E-21                 | <i>CADPS2</i>                             | Yes      | Yes      |
| 8105302 | 9.95               | 221.66      | 4.24E-21                 | <i>FST</i>                                | Yes      | Yes      |
| 8103415 | 0.30               | 220.28      | 4.39E-21                 | <i>C4orf18</i>                            | Yes      |          |
| 8180078 | 2.71               | 217.91      | 4.86E-21                 | <i>HLA-DMB</i>                            |          |          |
| 8151334 | 3.84               | 210.18      | 7.69E-21                 | <i>MSC</i>                                | Yes      |          |
| 8001564 | 6.22               | 206.03      | 9.73E-21                 | <i>DOK4</i>                               | Yes      | Yes      |
| 8054439 | 10.58              | 204.66      | 9.88E-21                 | <i>ST6GAL2</i>                            | Yes      | Yes      |
| 8097910 | 1.79               | 204.54      | 9.88E-21                 | <i>FGB</i>                                |          |          |
| 7969493 | 0.30               | 199.75      | 1.30E-20                 | <i>SCEL</i>                               | Yes      |          |

Of the 1,474 significantly differentially expressed genes from the IC25 dose 1 week recovery time course, 88 genes were found to have a greater than three-fold increase or decrease in expression on comparison with round 0 with the final round 15; 22 of which had greater than five-fold change. In contrast with the IC50 dose 3 week recovery time course, the majority of the genes showed at least a three-fold increase in expression (75 of the 88) between round 0 and round 15. Overall, 816 genes were found to have increased fold change expression on comparison between rounds 0 and 15, while 658 had decreased fold change over the same period.

Table 4.2 shows the top 25 differentially expressed genes (for a more comprehensive list [379 gene with B-H adjusted  $p$ -value less than or equal to  $1E-14$ ], see E-Appendix 9.1.2). The top ten genes from this time course appear to fall into more distinct categories than their counterparts from the IC50 dose 3 week recovery time course. One overlaps both top ten lists, *BAAT*, and only one other metabolism-related enzyme is in this top ten, fatty acid CoA reductase 2 (*FAR2*). Three members from the annexin family of calcium ion and phospholipid binding proteins (*ANXA8*, *ANXA8L1*, and *ANXA8L2*) show a slight increase in fold change from round 0 to round 15 (1.59, 2.55, and 1.60 respectively); their over expression has been linked to pancreatic ductal adenocarcinoma (Karanjawala *et al.*, 2008). Molecules involved in adhesion and the extracellular matrix can also be seen in these top genes (neuronal cell adhesion molecule, *NRCAM*; filamin A interacting protein 1, *FILIP1*; and collagen and calcium binding EGF domains 1, *CCBE1*). *CCBE1* is also down-regulated in ovarian cancer cell lines, suggesting a role for it in tumour suppression. It is conceivable that uroplakin 1B (*UPK1B*) and glypican 6 (*GPC6*) could also be having strong phenotypic effects in response to cisplatin treatment as a transmembrane protein and a proteoglycan implicated in cell growth and division (all <http://www.ncbi.nlm.nih.gov/gene>).

**Table 4.2** IC25 dose 1 week recovery; top 25 differentially expressed genes. Fold change (increase in green, decrease in red) shows change in expression when comparing the chemo-naïve (round 0) cells with those from the end of the time course (round 15).

| ID      | Fold Change Rd0v15 | F Statistic | B-H adj.p.Value | Gene Symbol    | > 3fold? | > 5fold? |
|---------|--------------------|-------------|-----------------|----------------|----------|----------|
| 8142270 | 4.07               | 482.70      | 3.59E-31        | <i>NRCAM</i>   | Yes      |          |
| 7927307 | 1.60               | 360.44      | 3.55E-29        | <i>ANXA8L2</i> |          |          |
| 8081826 | 0.90               | 349.15      | 4.20E-29        | <i>UPK1B</i>   |          |          |
| 7954631 | 0.23               | 334.83      | 6.71E-29        | <i>FAR2</i>    | Yes      |          |
| 7933312 | 2.55               | 316.32      | 1.34E-28        | <i>ANXA8L1</i> |          |          |
| 7927367 | 1.59               | 315.04      | 1.34E-28        | <i>ANXA8</i>   |          |          |
| 8162870 | 4.93               | 300.19      | 2.75E-28        | <i>BAAT</i>    | Yes      |          |
| 7969613 | 4.90               | 279.01      | 8.99E-28        | <i>GPC6</i>    | Yes      |          |
| 8127646 | 4.06               | 256.54      | 3.62E-27        | <i>FILIP1</i>  | Yes      |          |
| 8023575 | 1.69               | 233.05      | 1.83E-26        | <i>CCBE1</i>   |          |          |
| 7936494 | 1.13               | 224.94      | 3.14E-26        | <i>GFRA1</i>   |          |          |
| 8139212 | 0.14               | 216.02      | 5.94E-26        | <i>GLI3</i>    | Yes      | Yes      |
| 7927280 | 0.89               | 190.87      | 5.03E-25        | <i>PPYR1</i>   |          |          |
| 8152617 | 3.65               | 189.61      | 5.26E-25        | <i>HAS2</i>    | Yes      |          |
| 7965156 | 0.97               | 182.08      | 1.01E-24        | <i>LIN7A</i>   |          |          |
| 8114920 | 0.25               | 179.82      | 1.19E-24        | <i>DPYSL3</i>  | Yes      |          |
| 7957260 | 1.83               | 175.35      | 1.75E-24        | <i>GLIPR1</i>  |          |          |
| 8175710 | 0.79               | 172.86      | 2.13E-24        | <i>CSAG2</i>   |          |          |
| 8115327 | 6.10               | 169.99      | 2.72E-24        | <i>SPARC</i>   | Yes      | Yes      |
| 7977786 | 6.67               | 169.29      | 2.78E-24        | <i>SLC7A7</i>  | Yes      | Yes      |
| 7965206 | 0.66               | 168.80      | 2.79E-24        | <i>SLC6A15</i> |          |          |
| 8066214 | 3.71               | 166.96      | 3.24E-24        | <i>TGM2</i>    | Yes      |          |
| 8151334 | 2.21               | 163.49      | 4.51E-24        | <i>MSC</i>     |          |          |
| 8170580 | 0.81               | 155.05      | 1.11E-23        | <i>CSAG2</i>   |          |          |
| 7902565 | 4.18               | 152.71      | 1.39E-23        | <i>LPN2</i>    | Yes      |          |

Further down the list (for full listing see E-Appendix 9.1.2) are GLI family zinc finger 3 (*GLI3*), a DNA-binding transcriptional activator and repressor of Sonic hedgehog signalling (Chapter 1 Section 1.6.4), and *Chondrosarcoma-Associated Gene 2* (*CSAG2*), also known as the taxol-resistant-associated gene 3 protein, whose expression has been associated with chemotherapy resistance and neoplastic phenotypes (<http://www.genecards.org>). The presence of these genes and those from the IC25 dose 1 week recovery top ten differentially expressed genes and their more obvious links to hallmarks of cancer or potential drug resistance may indicate that the transcriptomes between the cells from the two time courses are quite different as a result of their divergent treatment courses.

#### **4.4 Temporal Cluster Analysis**

Transcripts were clustered according to their expression profiles, allowing transcripts with similar increases or decreases in expression to be visualised, potentially collecting the transcripts that may be working together. This was done using the fuzzy *c*-means algorithm to soft cluster the members (Kumar and E Futschik, 2007), which is a noise-robust cluster algorithm that can be used for time-series data. This is less stringent than the hard *k*-means clustering and allows smaller changes to be detected. Each point is associated with a cluster dependent on a membership value (*m*), which was chosen as 1.25 for all instances. The value of *m* is used to dictate how strongly each gene belongs to a cluster. Using *m*=1.25 ensures all genes are strongly associated with a particular cluster without removing flexibility. It is less stringent than *k*-means-like clustering ( $m \leq 1.1$ ) but not as dispersed as  $m \geq 1.6$  (Dembélé and Kastner, 2003). Expression for each gene was standardised so that its average expression is zero with a standard deviation of one.

It is standard procedure to perform clustering multiple times, varying the number of clusters and looking to find a satisfactory compromise between the number of genes per cluster, the number of clusters, and the observed temporal changes in expression. Overlap plots were used to help decide this with Principle Components Analysis (PCA) performed on each cluster to investigate the similarities and differences between them as well as allowing their graphically presentation (Section 4.4.1).

Examination of the patterns of expression over the course of each time course identified clusters that either increased or decreased overall during cisplatin treatment. These clusters may be considered most likely to be biologically relevant and so were submitted to in-depth analysis by Gene Ontology (GO). Genes found within these increasing and decreasing expression clusters with highly significant differential expression ( $p$ -value less than  $1E-14$ ) were reviewed in the published literature.

GO term enrichment analysis and functional annotations were generated using the Database for Annotation, Visualization and Integrated Discovery (DAVID) (Chapter 2 Section 2.2.3.5). The GO initiative aims to bring together the representation of gene and gene product attributes. It categorises genes into one of three main domains; cellular components, molecular functions, and biological processes. DAVID aims to provide functional interpretation and enhance the context of large gene lists.

#### 4.4.1 Expression Clustering

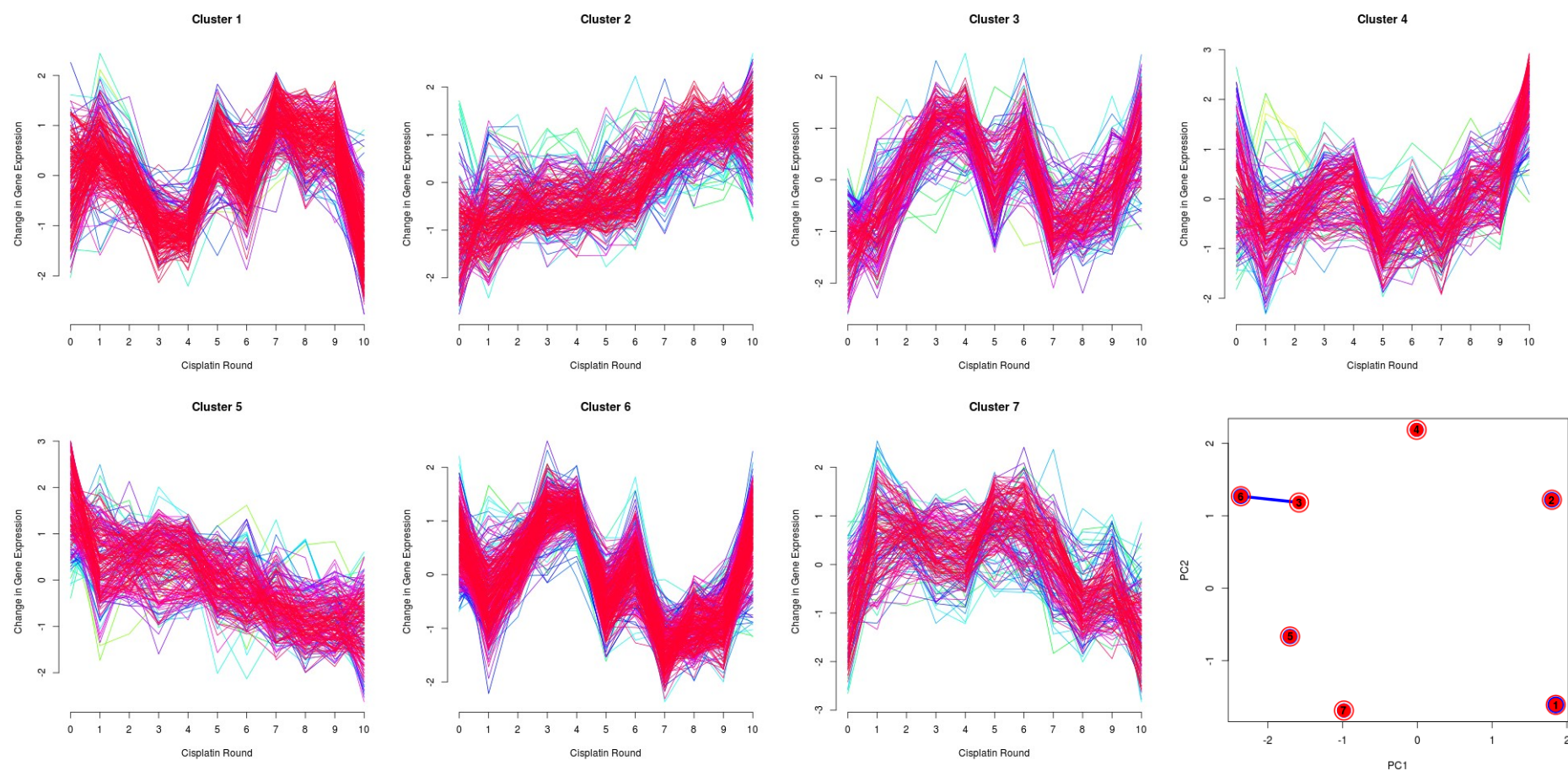
##### *4.4.1.1 IC50 Dose 3 Week Recovery Time Course*

Temporal expression clustering was performed on the top 1,515 genes with a B-H adjusted  $p$ -value of less than or equal to  $1\text{E-}08$  for significance in change in gene expression over the time course. After multiple iterations of cluster numbers, seven modules of genes were chosen that showed distinctive temporal patterns, characterised by their initial levels of abundance and variation in expression over cisplatin treatments (Table 4.3 and Figure 4.7).

| Cluster      | 1   | 2   | 3   | 4   | 5   | 6   | 7   |
|--------------|-----|-----|-----|-----|-----|-----|-----|
| No. of Genes | 230 | 227 | 187 | 167 | 209 | 325 | 170 |

**Table 4.3** Number of genes within each Expression Cluster for the IC50 dose with 3 week recovery time course.

**Figure 4.7** IC50 Dosing with 3 Week Recovery: fuzzy *c*-means cluster analysis. The change in log2 fold expression of genes within each cluster are shown over the course of the cisplatin rounds, highlighting some of the large fold changes seen over the course of treatment. The depth of red colour shows how strongly each gene is associated with a particular cluster. The bottom right overlap plot shows the PCA between the clusters; the blue line shows the similarity between clusters 3 and 6, however the two start at different levels so the two clusters were not combined.



Clusters 2 and 5 are of particular interest because they exhibit either a general increase (Cluster 2) or a general decrease (Cluster 5) over the course of cisplatin dosing. GO analysis for these two clusters was conducted using DAVID, to investigate potential signalling pathways (as detailed in Chapter 2 Section 2.2.3.5).

For Cluster 2, the most enriched functional annotation (enrichment score of 4.73 [3 significant figures (sf)]) contained mainly signalling peptides and glycoprotein terms (up to 37% of the members of the expression cluster), and genes encoding proteins that are found in the extracellular region (B-H adjusted  $p$ -value 3.26E-05 [3 sf]). The second most enriched functional annotation cluster (enrichment score of 2.44 [3 sf]) were EGF and EGF-like domains (up to 6% of the number of molecules in the expression cluster) (B-H adjusted  $p$ -value 1.17E-03 [3 sf]). These two functional annotation groups suggest these up-regulated genes are having effects on the signalling cascades within the cell.

Similar functions to those of Cluster 2 were observed for Cluster 5. The two most enriched functional annotation clusters (enrichment score of 5.93 and 4.97 [3 sf]) contained signalling peptides and glycoprotein terms (up to 36% of the number of molecules in the expression cluster; B-H adjusted  $p$ -value 1.44E-05 [3 sf]) or were proteins in the extracellular region (up to 31% of the number of molecules in the expression cluster). Functional annotation group 3 (enrichment score 4.38) constituted responses to various stimuli or substances, such as organic substances (13% of number of molecules in the expression cluster; B-H adjusted  $p$ -value 1.29E-04) or endogenous stimulus (9% of number of molecules in the expression cluster; B-H adjusted  $p$ -value 8.18E-05). Another significant functional annotation suggested the enrichment of adhesion molecules within this cluster, with terms such as basolateral plasma membrane (6% of number of molecules in the expression cluster; B-H adjusted  $p$ -

value 1.40E-04) and anchoring junction (5% of number of molecules in the expression cluster; B-H adjusted  $p$ -value 5.48E-04) appear as the most significant. A full listing of the genes in Clusters 2 and 5 is available in E-Appendix 9.1.3.

The genes showing the most significant differential expression from these two clusters of interest are discussed later in Section 4.6.

#### 4.4.1.2 IC25 Dose 1 Week Recovery Time Course

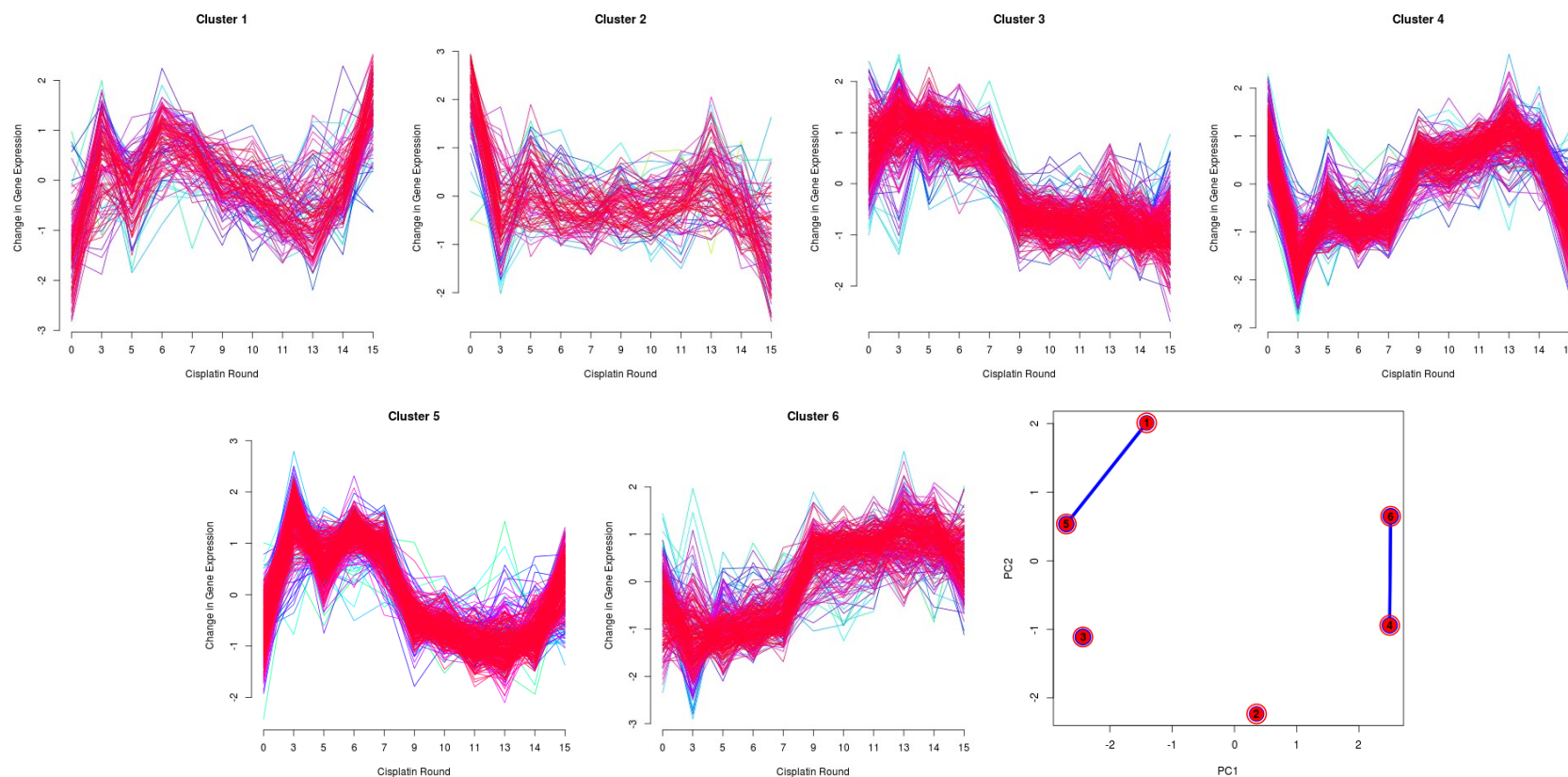
Temporal expression clustering was performed on the top 1,474 genes with a B-H adjusted  $p$ -value of less than or equal to 1E-08 for significance in change in gene expression over the time course. After multiple iterations of cluster numbers, six modules of genes were chosen that showed distinctive temporal patterns, characterised by their initial levels of abundance and variation in expression over cisplatin treatments (Table 4.4 and Figure 4.8).

| Cluster     | 1   | 2   | 3   | 4   | 5   | 6   |
|-------------|-----|-----|-----|-----|-----|-----|
| No of Genes | 118 | 103 | 323 | 286 | 359 | 285 |

**Table 4.4** Number of genes within each Expression Cluster for the IC25 dose with 1 week recovery time course.

Clusters 3 and 6 are of particular interest due to the patterns they exhibit over the course of cisplatin dosing – general increase (Cluster 6) and general decrease (Cluster 3). For these two clusters, GO analysis was conducted using DAVID to investigate potential signalling pathways as detailed in Chapter 2 Section 2.1.4.5.

**Figure 4.8** IC25 Dose with 1 Week Recovery: fuzzy *c*-means cluster analysis. The change in log2 fold expression of genes within each cluster are shown over the course of the cisplatin rounds, highlighting some of the large fold changes seen over the course of treatment. The depth of red colour shows how strongly each gene is associated with a particular cluster. The far bottom right overlap plot shows the PCA between the clusters. The blue line on the PCA plot show the similarities between clusters 1 and 5, and between clusters 4 and 6, but using four clusters instead of six would give an unacceptably high number of genes per cluster. There are also significant differences in levels of log2 change in expression across the time course.



Cluster 6 showed high enrichment in GO terms that are involved in protein dimerization, glycoproteins and signal peptides (up to 31% of molecules in the annotation cluster), and response to environmental stimuli, such as hormones or organic compounds (3.47, 2.91, and 2.85 respectively) in spite of *p*-values only approaching significance. Regulation of EMT, morphogenesis, differentiation, and development also had a strong enrichment score (2.52) and, in combination with other GO terms enriched in this cluster, is suggestive of highly active cells responding to their environment.

For cluster 3, functional annotation clusters included a significantly high proportion of glycoproteins, including secreted signalling peptides (enrichment score 6.82; 14-38% of the molecules, B-H adjusted *p*-value 9.35E-08). Transmembrane proteins were also enriched significantly (enrichment score 3.57, B-H adjusted *p*-value 5.12E-06), as were immunoglobulins (enrichment score 2.97; B-H adjusted *p*-value 9.0E-03). A full listing of the genes in Clusters 3 and 6, including their B-H adjusted *p*-values, is available in E-Appendix 9.1.4.

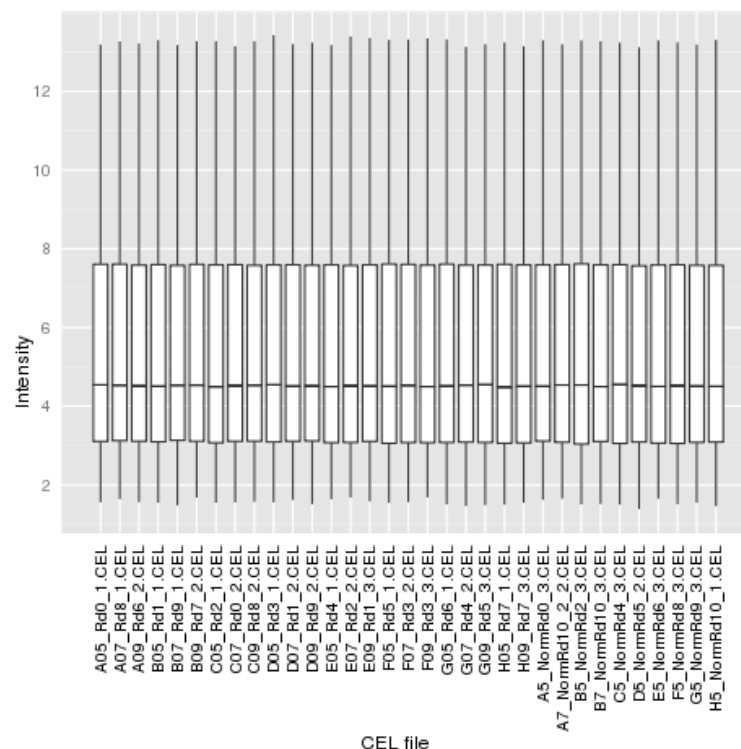
The genes showing the most significant differential expression from these two clusters of interest are discussed later in Section 4.6.

For all the up- and down-regulated gene clusters (2 and 5 in IC50 dose 3 week recovery; 3 and 6 in IC25 dose 1 week recovery), it is noteworthy that a slight sigmoidal curve is visible with the alteration in expression change occurring between rounds four and six of the IC50 dose 3 week recovery time course and between rounds seven and nine of the IC25 Dose 1 week recovery time course. It is possible that this small change reflects the acquisition of stronger resistance to cisplatin around this point in the time course as the exposure to the drug begins to take its effects on cells.

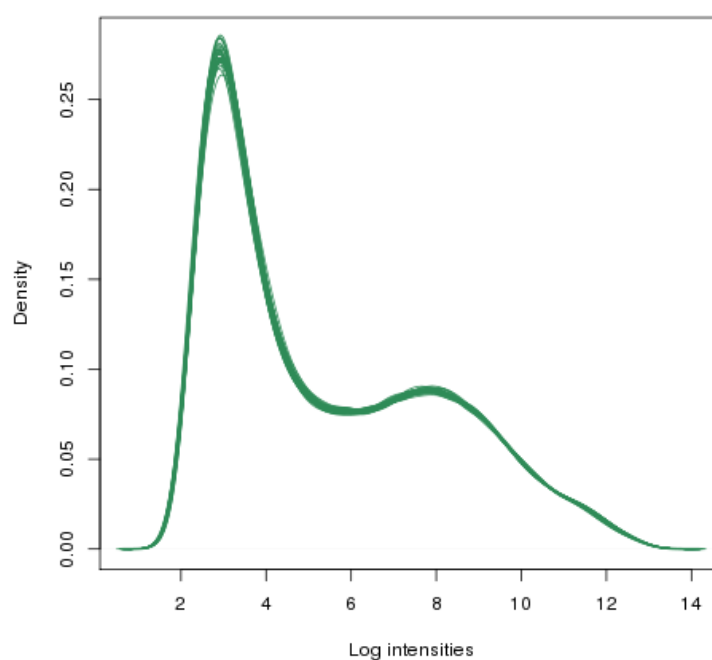
## 4.5 Gene Expression Analysis of Untreated A549 Cells

It is feasible that some of the changes in global gene expression observed for the two time courses may be due to repeated passages of the A549 cells in culture over a prolonged period (Chapter 3 Section 3.5). To control for this and to establish the changing genomics of A549 cells over numerous rounds of normal cell culture a time control was set up.

Cells were grown using the same initial seeding density ( $2 \times 10^6$  cells) as the IC50 dose 3 week recovery time course and RNA and DNA were harvested at the same equivalent time points and passages (Chapter 2 Section 2.2.1). RNA extraction, processing, and microarray hybridisation were performed as detailed previously (Chapter 2 Sections 2.1.4.1 and 2.1.4.2). The same microarray quality control measures were implemented (Chapter 2 Section 2.1.4.3) and no outliers were identified across the 33 microarrays (Figures 4.9 and 4.10).



**Figure 4.9** Log Expression Signal of Normal Time Course showing that all the box and whisker plots for the 33 samples from round 0 to 10 have a similar median, interquartile range and total range.



**Figure 4.10** Signal Density Distribution showing practically identical distributions across all the 33 arrays.

#### 4.5.1 Comparisons of the Significantly Differentially Expressed Genes between the Untreated and Treated Time Courses

Differential expression analysis was performed for the normal cell culture as outlined in Section 4.3 and the list of genes obtained was compared with those from the analyses of the IC50 dose 3 week recovery and IC25 dose 1 week recovery time courses. Using the same threshold of B-H adjusted  $p$ -value less than or equal to  $1E-08$ , resulted in 865 differentially expressed genes for the normal cell culture time course, substantially fewer than that obtained for either of the cisplatin-treated time courses. Overall, 338 genes were found to have increased fold change expression on comparison between rounds 0 and 10, while 527 had decreased fold change over the same period.

**Table 4.5** Normal untreated time course; top 25 differentially expressed genes. Fold change (increases in green, decreases in red) shows change in expression when comparing the chemo-naïve (round 0) cells with those from the end of the time course (round 10).

| ID      | Fold Change<br>Rd0v10 | F Statistic | B-H<br>adj.p.Value | Gene Symbol     | > 3fold? | > 5fold? | In IC50 dose 3 week<br>recovery? | In IC25 dose 1 week<br>recovery? |
|---------|-----------------------|-------------|--------------------|-----------------|----------|----------|----------------------------------|----------------------------------|
| 8112980 | 0.02                  | 485.87      | 9.18E-22           | <i>EDIL3</i>    | Yes      | Yes      | Yes                              | Yes                              |
| 8172022 | 0.06                  | 446.46      | 1.33E-21           | <i>TMEM47</i>   | Yes      | Yes      | Yes                              | Yes                              |
| 8095303 | 0.03                  | 426.94      | 1.55E-21           | <i>LPHN3</i>    | Yes      | Yes      | Yes                              | Yes                              |
| 8128284 | 0.09                  | 414.08      | 1.70E-21           | <i>EPHA7</i>    | Yes      | Yes      | Yes                              |                                  |
| 8180234 | 35.31                 | 393.78      | 2.56E-21           | <i>unknown</i>  | Yes      | Yes      | Yes                              |                                  |
| 8099633 | 0.08                  | 370.12      | 4.64E-21           | <i>PPARGC1A</i> | Yes      | Yes      | Yes                              | Yes                              |
| 8035095 | 27.03                 | 315.48      | 2.94E-20           | <i>CYP4F11</i>  | Yes      | Yes      | Yes                              | Yes                              |
| 8081036 | 0.07                  | 304.00      | 4.08E-20           | <i>CADM2</i>    | Yes      | Yes      | Yes                              | Yes                              |
| 8114920 | 3.16                  | 262.88      | 2.23E-19           | <i>DPYSL3</i>   | Yes      |          | Yes                              | Yes                              |
| 8036079 | 0.18                  | 231.59      | 9.74E-19           | <i>DMKN</i>     | Yes      | Yes      |                                  | Yes                              |
| 8142687 | 0.15                  | 228.19      | 1.06E-18           | <i>GPR37</i>    | Yes      | Yes      | Yes                              | Yes                              |
| 7927280 | 11.54                 | 224.99      | 1.16E-18           | <i>PPYR1</i>    | Yes      | Yes      | Yes                              | Yes                              |
| 8112731 | 10.88                 | 222.56      | 1.23E-18           | <i>F2RL2</i>    | Yes      | Yes      | Yes                              | Yes                              |
| 8148040 | 0.02                  | 217.69      | 1.47E-18           | <i>MAL2</i>     | Yes      | Yes      | Yes                              | Yes                              |
| 8115327 | 8.48                  | 216.85      | 1.47E-18           | <i>SPARC</i>    | Yes      | Yes      | Yes                              | Yes                              |
| 8090180 | 0.37                  | 213.12      | 1.71E-18           | <i>MUC13</i>    |          |          | Yes                              | Yes                              |
| 7957452 | 0.30                  | 209.98      | 1.94E-18           | <i>ALX1</i>     | Yes      |          | Yes                              | Yes                              |
| 8100541 | 5.21                  | 208.02      | 2.06E-18           | <i>IGFBP7</i>   | Yes      | Yes      | Yes                              | Yes                              |
| 8005048 | 7.02                  | 204.04      | 2.48E-18           | <i>MYOCD</i>    | Yes      | Yes      | Yes                              |                                  |
| 8136662 | 0.27                  | 201.17      | 2.70E-18           | <i>MGAM</i>     | Yes      |          | Yes                              | Yes                              |
| 7978801 | 0.10                  | 201.00      | 2.70E-18           | <i>MDGA2</i>    | Yes      | Yes      |                                  |                                  |
| 8134098 | 0.16                  | 196.28      | 3.47E-18           | <i>PFTK1</i>    | Yes      | Yes      | Yes                              |                                  |
| 8160546 | 0.32                  | 194.40      | 3.74E-18           | <i>LINGO2</i>   | Yes      |          | Yes                              | Yes                              |
| 8155802 | 0.27                  | 191.85      | 4.22E-18           | <i>GDA</i>      | Yes      |          | Yes                              | Yes                              |
| 8103415 | 0.24                  | 188.89      | 4.92E-18           | <i>C4orf18</i>  | Yes      |          | Yes                              | Yes                              |

Table 4.5 shows the top 25 differentially expressed genes from the normal untreated time course. It is noteworthy that only one of these genes, *MDGA2*, is not found to be differentially expressed in either of the two treated time courses.

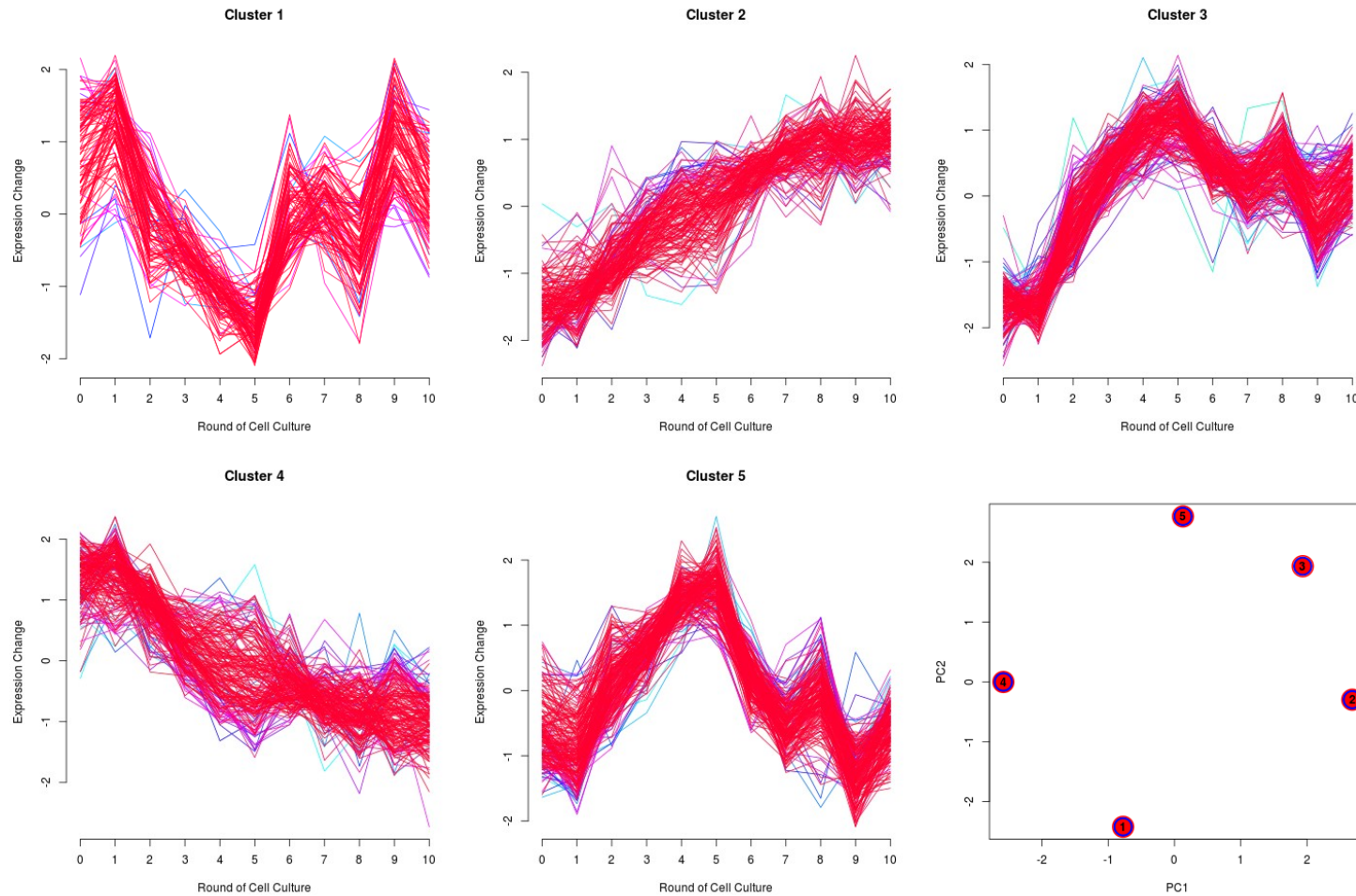
The overlap of genes between the treated and untreated time courses, in general, was found to be quite high: 628 out of 1515 (41.5%) normal versus IC50 dose 3 week recovery and 560 out of 1474 (38.0%) normal versus IC25 dose 1 week recovery. It would be desirable to know if these significantly differentially expressed genes that overlap all three time courses (normal untreated, IC50 dose 3 week recovery, and IC25 1 week recovery) have the same expression profiles. To investigate this, fuzzy *c*-means clustering (Section 4.4) was performed on the 865 genes found to be significantly differentially expressed across the untreated time course. After testing several different numbers of clusters, five were chosen as each showed a unique expression patterns over the course of culture and contained a manageable number of genes (Table 4.6).

| Cluster     | 1   | 2   | 3   | 4   | 5   |
|-------------|-----|-----|-----|-----|-----|
| No of Genes | 114 | 160 | 208 | 183 | 200 |

**Table 4.6** Number of genes within each temporal expression cluster for the 865 differentially expressed genes from the untreated normal time course.

The expression patterns of the five clusters (Figure 4.11) exhibited similar patterns to the clusters observed for the two cisplatin time courses (Figures 4.7 and 4.8). Clusters showing a gradual increase (Cluster 2) and a gradual decrease (Cluster 4) over the time course are present. Genes within these clusters with increasing or decreasing expression are of particular interest since if they have the same expression profile in the equivalent clusters in the treated time courses, these could be confounders when searching for potential biomarkers of cisplatin resistance.

**Figure 4.11** Untreated Normal Time Course: fuzzy *c*-means cluster analysis. Each time point was matched as closely as possible to those from the treated time courses. The log<sub>2</sub> fold expression change of the 865 differentially expressed genes across each cluster are shown over the “rounds”. The bottom right overlap plot shows the PCA between the clusters; each cluster is distinct, as indicated by five discrete expression profiles.



GO analyses of clusters 2 and 4 showed that for both clusters the most enriched functional annotation clusters were glycoproteins (4.95 enrichment score, 66 count, 1.2E-05 B-H adjusted *p*-value; 4.87 enrichment score, 70 count, 7.7E-05 B-H adjusted *p*-value respectively) closely followed by plasma membrane fraction terms, suggesting that signalling again may be the main focus of the normal time course. The absence of significant terms for the processes previously seen in the cisplatin-treated time course clusters (Section 4.4.1), including the response to stimuli or regulation of key cell processes such as differentiation, highlights that these pathways may be of importance in the response to cisplatin.

Comparing the gene contents of each of these clusters with the comparable ones from the cisplatin time courses only a small proportion, generally less than ten per cent, were found to overlap in each of the respective clusters (Table 4.7 and Table 4.8). Nevertheless, since the effects of these genes may well confound genuine results or potential biomarkers of changes caused by cisplatin, these genes were excluded from further analyses.

|                                                   | Up in Normal Cells<br>(Cluster 2) |                                                     | Down in Normal Cells<br>(Cluster 4) |
|---------------------------------------------------|-----------------------------------|-----------------------------------------------------|-------------------------------------|
| Up in IC50 Dose 3<br>Week Recovery<br>(Cluster 2) | 11 / 227                          | Down in IC50 Dose 3<br>Week Recovery<br>(Cluster 5) | 14 / 209                            |
| Up in IC25 Dose 1<br>Week Recovery<br>(Cluster 6) | 27 / 285                          | Down in IC25 Dose 1<br>Week Recovery<br>(Cluster 3) | 22 / 323                            |

**Table 4.7** Comparison of the number of differentially expressed genes within each increasing or decreasing expression cluster with the up-regulated and down-regulated gene clusters in the untreated normal time course. **Table 4.8** shows the genes that overlap these clusters.

| Treated Time Course        | Increased Gene Expression in both the Treated and Untreated Time Courses                                                                                                                                                              | Decreased Gene Expression in both the Treated and Untreated Time Courses                                                                                                                                       |
|----------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| IC50 dose 3 weeks recovery | <i>LUM, CEACAM5, NTRK3, <b>TNFSF10</b>, RSPO3, ADAM28, THBD, PCOLCE2, <b>ZNF160</b>, PREX1, MAMLD1</i>                                                                                                                                | <i><b>FXYD2</b>, SNAI2, EHF, TSPAN12, RDH10, TF, <b>IGFBP7</b>, NKD1, <b>FJX1</b>, DCBLD1, SORT1, LIMA1, <b>GNAL</b>, FAM13A</i>                                                                               |
| IC25 dose 1 week recovery  | <i>HAS2, LPHN2, MOBKL2B, RP11-35N6.1, HGD, BCAS1, TM4SF20, LHFP, LOC554202, THY1, FRMD5, CNTNAP3, SERPINE2, CSF1, RASSF5, CIQTNF1, PBLD, <b>TNFSF10</b>, AKAP12, <b>ZNF160</b>, MYO15B, EFHD2, LAMB3, CREB5, THBD, GRB10, ST3GAL1</i> | <i>RNF182, SERPINE1, TINAG, SPOCK1, LYPD1, <b>IGFBP7</b>, ANKS4B, LGALS2, ADORA2B, TIPARP, CEACAM6, <b>FXYD2</b>, <b>FJX1</b>, P4HA3, CYP4F3, LGALS3BP, MUC5B, GABARAPL1, CFI, TTC39A, <b>GNAL</b>, TMEM47</i> |

**Table 4.8** 68 unique genes that share similar increasing and decreasing expression profiles over the course of cisplatin treatment and extended untreated cell culture rounds. Those genes in **bold** highlight those few genes that are seen to have the same expression profiles across both treated time courses as well as in the untreated. All genes within this table were excluded at this stage from future investigations.

The combination of the relatively high overlap of differentially expressed genes (41% and 38% Section 4.5.1) but the lack of overlap in temporal expression seen here is interesting. It is possible that the effects of cisplatin treatment are altering the normal processes of the cells to enable them to continue to survive in the presence of the drug.

Comparing the untreated Clusters 2 and 4 with their equivalent clusters from the cisplatin time courses, it may be of interest that the up- and down-regulated clusters appear to have a more continuous increase or decrease in their expression profiles than their respective counterparts in Figures 4.7 and 4.8; no sigmoidal shape is visible in either.

#### 4.5.2 Investigating the Overlap between the Time Courses

Examination of the two gene lists, 1,515 genes for the IC50 dose 3 week recovery time course and 1,474 genes for the IC25 dose 1 week recovery time course revealed that 843 genes were common to both time courses. It is likely that these genes are involved in mechanisms of resistance irrespective of the dosing regimen or the continuing propagation of the cell line over time. The 843 genes in common accounted for 55.6% and 57.2% of the IC50 dose 3 week recovery and IC25 dose 1 week recovery genes respectively (Figure 4.12).



**Figure 4.12** Venn diagram showing the overlap of 843 gene between the those that are most significantly differentially expressed in the IC50 dose 3 week recovery time course and the IC25 dose 1 week recovery time course.

Three hundred and 75 of the 843 overlapping genes were found to be only present in the treated time courses and not in the untreated normal time course (threshold of adjusted B-H  $p$ -value of less than or equal to  $1E-08$ ). These 375 genes are of particular interest. The 28

genes with the largest fold changes when comparing round 0 with round 10 from the IC50 dose 3 week recovery time course can be found in Table 4.9. A full list of all 375 genes can be found in E-Appendix 9.1.5.

Two of these genes show especially large fold changes; ankyrin repeat domain 30A (*ANKRD30A*) had an 11.9-fold increase while tolloid-like1 (*TLL1*) exhibited a 12.5-fold decrease. Also in these genes is lysophosphatidic acid receptor 5 (*LPAR5*), a gene that encodes a member of the rhodopsin class of G protein-coupled transmembrane receptors that are upstream regulators of Rho and Rac GTPase pathways (Fleming *et al.*, 1997). Antagonism of the receptor family has been shown to cause breast cancer tumour regression *in vivo* (Zhang *et al.*, 2009). Other genes associated with GTPase pathways are also found in this list, such as TBC1 domain family member 2 (*TBC1D2*) that acts as a GTPase activating protein for RAB7A and a signal effector acting as a linker between RAC1 and RAB7A, eventually leading to reduced cell-cell adhesion (<http://www.genecards.org>). Galanin-GMAP prepropeptide (GAL) is a small neuropeptide that functions as cellular messenger in central and peripheral nervous system. It and its receptor GAI1 have been identified as novel determinants of chemotherapy resistance and potential biomarkers in colorectal cancer (Stevenson *et al.*, 2012).

**Table 4.9** The 28 from the 375 overlapping differentially expressed genes between the two treated time courses with at least a three-fold change in expression between the start (round 0) and end (round 10) using data from the IC50 dose 3 week recovery time course. Fold change is represented with green and red indicating an increase and decrease respectively.

| ID      | Fold Change Rd0v10 | F Statistic | B-H adj.p.value | Gene Symbol                 | > 3fold? | > 5fold? |
|---------|--------------------|-------------|-----------------|-----------------------------|----------|----------|
| 8068202 | 0.13               | 398.05      | 3.85E-24        | <i>C21orf63</i>             | Yes      | Yes      |
| 8098214 | 0.08               | 312.96      | 7.29E-23        | <i>TLL1</i>                 | Yes      | Yes      |
| 7927033 | 11.87              | 160.44      | 1.36E-19        | <i>ANKRD30A</i>             | Yes      | Yes      |
| 8023727 | 0.20               | 121.10      | 3.76E-18        | <i>DSEL</i>                 | Yes      |          |
| 8162759 | 0.19               | 104.47      | 2.25E-17        | <i>TBC1D2</i>               | Yes      | Yes      |
| 7974835 | 0.26               | 100.73      | 3.35E-17        | <i>PRKCH</i>                | Yes      |          |
| 7927058 | 3.78               | 67.24       | 4.74E-15        | Transcript cluster on Chr10 | Yes      |          |
| 8009243 | 0.23               | 66.51       | 5.29E-15        | <i>C17orf60</i>             | Yes      |          |
| 8109490 | 4.55               | 65.85       | 5.95E-15        | <i>SGCD</i>                 | Yes      |          |
| 7946067 | 5.50               | 62.42       | 1.13E-14        | <i>OR51B4</i>               | Yes      | Yes      |
| 7946061 | 6.05               | 59.76       | 1.90E-14        | <i>HBE1</i>                 | Yes      | Yes      |
| 7960637 | 3.14               | 57.29       | 3.09E-14        | <i>LPAR5</i>                | Yes      |          |
| 8101762 | 0.31               | 57.24       | 3.12E-14        | <i>SNCA</i>                 | Yes      |          |
| 7947282 | 4.05               | 56.52       | 3.64E-14        | <i>DCDC5</i>                | Yes      |          |
| 8037750 | 0.24               | 54.97       | 5.07E-14        | <i>IGFL4</i>                | Yes      |          |
| 8088560 | 0.30               | 54.28       | 5.84E-14        | <i>ADAMTS9</i>              | Yes      |          |
| 8059680 | 3.50               | 51.89       | 9.95E-14        | <i>HTR2B</i>                | Yes      |          |
| 8124351 | 5.57               | 49.95       | 1.56E-13        | <i>SLC17A3</i>              | Yes      | Yes      |
| 7901175 | 0.32               | 43.54       | 8.13E-13        | <i>TSPAN1</i>               | Yes      |          |
| 7942064 | 4.59               | 42.49       | 1.08E-12        | <i>GAL</i>                  | Yes      |          |
| 8081488 | 3.56               | 41.40       | 1.49E-12        | <i>HHLA2</i>                | Yes      |          |
| 8062813 | 3.12               | 39.34       | 2.69E-12        | <i>R3HDML</i>               | Yes      |          |
| 7903227 | 0.23               | 38.60       | 3.38E-12        | <i>PALMD</i>                | Yes      |          |
| 7956867 | 3.85               | 35.15       | 9.98E-12        | <i>HMGA2</i>                | Yes      |          |
| 8151496 | 0.32               | 29.16       | 8.48E-11        | <i>ZNF704</i>               | Yes      |          |
| 8046078 | 0.33               | 23.34       | 1.04E-09        | <i>B3GALT1</i>              | Yes      |          |
| 8021695 | 0.32               | 19.64       | 6.88E-09        | <i>DOK6</i>                 | Yes      |          |
| 8146633 | 0.29               | 19.53       | 7.34E-09        | <i>LOC643763</i>            | Yes      |          |

Further literature-based investigations into the full list of 375 genes (E-Appendix 9.1.5) have also yielded some more remarkable results. Known oncogenes and proto-oncogenes, including *RAB26*, adreno-receptor alpha 1B (*ADRA1B*), *Jun B proto-oncogene (JUNB)*, and *RAB38*, are present as are genes known to have influences on important signalling pathways in cancer, including the EGFR and KRAS pathways. In addition the WNT beta-catenin pathway seems to be playing a role in these cisplatin-treated cells as genes encoding several regulators, ligands and receptors of this signalling cascade (necleoredoxin, *NXN*; secreted frizzled-related protein 4, *SFRP4*; dickkopf 1 homolog, *DKK1*; frizzled family receptor 5, *FZD5*) are contained within the gene list (E-Appendix 9.1.5).

Important regulators of the cell cycle are found within these 375 overlapping genes, such as Cyclin D3 (*CCND3*), which may indicate the cell is attempting to prevent the cells, damaged by cisplatin exposure, from replicating. The presence of these effectors is reinforced by a recent publication revealing that cisplatin-resistant NSCLC cells accumulate in the G0/G1 phase of the cell cycle (Barr *et al.*, 2013). The incidence of many candidate tumour suppressor genes including platelet-derived growth factor receptor-like (*PDGFRL*), *H19*, protein tyrosine phosphatase receptor type G (*PTPRG*), c-MYC binding protein bridging integrator 1 (*BINI*), and patched 1 (*PTCH1*) also supports the idea that cell cycle control is being influenced by the presence of cisplatin.

In agreement with the morphological changes that were observed in the cells undergoing cisplatin treatment (Chapter 3 Section 3.3.), several genes involved in cell adhesion, EMT, and junctions between cells, including *NRCAM*, L1 cell adhesion molecule (*LICAM*), sidekick cell adhesion molecular 1 (*SDK1*), cadherin 12 type 2 (*CDH12*, Table 4.10), and components of collagen molecules, were found within the 375 genes significantly differentially expressed in the two treated time courses.

Within these 375 overlapping genes are several genes that overlap the same expression profile clusters (Sections 4.4.1.1 and 4.4.1.2). Twenty four genes appear in the two increasing clusters (Cluster 2 IC50 dose 3 week recovery time course and Cluster 6 IC25 dose 1 week recovery time course), while 21 genes are found in both decreasing clusters (Cluster 5 IC50 dose 3 week recovery time course and Cluster 3 IC25 dose 1 week recovery time course). From Table 4.10 it can be seen that some of genes from this list of 45 have been implicated in cancer before, such as the *EGFR* family member, v-erb-b2 erythroblastic leukemia viral oncogene homolog 3 (*ERBB3* [Young *et al.*, 2013]), the cell cycle protein Cyclin G2 (*CCNG2* [Zimmermann *et al.*, 2012]), the apoptosis regulator Fas cell surface death receptor (*FAS* [Wu *et al.*, 2013]), and TOX high mobility group box family member 3 (*TOX3* [Dittmer *et al.*, 2011]), which has known genetic variants that are associated with breast cancer.

| Genes in both Increasing Expression Clusters                                                                                                                        | Genes in both Decreasing Expression Clusters                                                                                                            |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>TOX3, ABCG2, TNC, NOTCH3, SCNN1A, HOXB2, LSR, ETV4, CCNG2, MCAM, FAM46A, GATSL1, MPZL3, ICA1, EBF1, COL5A1, DPYSL5, NR5A2, SFRP1, DUSP6, BACE2, MEIS2, PLCE1</i> | <i>TSPAN8, C17orf60, C7orf68, TNS4, PAQR5, ERBB3, ACOX2, HHLA2, MAML3, GATM, PTPN14, HAPLN3, GJA1, CDH12, CXXC5, RIC3, FAS, ZYX, TLE4, UHRF1, FABP5</i> |

**Table 4.10** Genes overlapping the two increasing expression profile clusters (Cluster 2 IC50 dose 3 week recovery time course and Cluster 6 IC25 dose 1 week recovery time course) and the two decreasing expression profile clusters (Cluster 5 IC50 dose 3 week recovery time course and Cluster 3 IC25 dose 1 week recovery time course).

GO analyses using DAVID of the 375 genes reinforce the findings from the individual genes. The most heavily enriched terms were “signal” and “signal peptide” (enrichment score 6.56; B-H adjusted *p*-value 4.31E-07; 107 genes), closely followed by response to stimuli, such as organic substance (GO:0010033), endogenous stimulus (GO:0009719), and steroid hormone stimulus (GO:0048545) (enrichment score 3.72; B-H adjusted *p*-value 0.019; 68 genes total). Cell adhesion also featured highly in the GO analysis (enrichment score 3.54; B-H adjusted

*p*-value 0.0019; 23 genes) and the epidermal growth factor receptor signalling pathway (GO:0007173) was found to be enriched (enrichment score 3.17). Other enriched annotation clusters also included heparin binding (enrichment score 2.85; B-H adjusted *p*-value 0.075) and regulation of cell size (enrichment score 2.51; B-H adjusted *p*-value 0.051). All these annotation clusters suggest that the cells are in a highly active state and are under positive growth signals.

## 4.6 Discussion

Investigating the differential gene expression profiles across the course of cisplatin treatment and continued cell culture has yielded some fascinating results. Temporal expression profiling of the time courses has given clusters of gene expression that can be used for identifying important candidates that may be involved in the mechanisms of cisplatin resistance. These clusters and the genes within in them may help elucidate ways the cells are evading death and continue to grow.

Those genes that have increasing expression over the course of cisplatin treatment (IC50 dose 3 week recovery cluster 2; IC25 dose 1 week recovery cluster 6; Section 4.4.1; E-Appendices 9.1.3 and 9.1.4) may be considered as likely to contain genes that:

- Promote proliferation in spite of the presence of cisplatin.
- Enable cisplatin to be removed from the cells via transport.
- Help the cell to avoid programmed cell death.
- Enable damaged DNA to be repaired and allow continued cell survival and proliferation.
- Promote an invasive cell phenotype.

- Enhance replicative potential.

In contrast, those genes whose expression decreases over the course of cisplatin treatment (IC50 dose 3 week recovery cluster 5; IC25 dose 1 week recovery cluster 3; Section 4.4.1; E-Appendices 9.1.3 and 9.1.4) may contain candidate genes that:

- Act as tumour suppressors.
- Promote cell death and apoptosis.
- Cause cell cycle arrest.
- Promote senescence of cells.

Examples of genes for these processes have been seen in the clusters of interest, as have many that have already been associated with cancers, including NSCLC and others such as ovarian (Materna *et al.*, 2007), breast (Shehata *et al.*, 2008, Natarajan *et al.*, 2012, Sukowati *et al.*, 2012), prostate (Shehata *et al.*, 2008), hepatocellular (Sun *et al.*, 2010, Zhang *et al.*, 2013b), and colorectal (Zhang *et al.*, 2013b) tumours.

When examining the top differentially expressed genes (B-H adjusted *p*-value less than 1E-14) showing increasing expression over the IC50 dose 3 week recovery time course (Cluster 2, Section 4.4.1.1) excluding those named in Table 4.8, some candidates for further investigations appear, such as the transporter ATP-binding cassette sub-family G member 2 (ABCG2, Table 4.10). ABCG2 is also known as the breast cancer resistance protein (BCRP) and is thought to be involved in the transport and efflux of drugs from cells (Sukowati *et al.*, 2012). MAL2 (Section 4.3.1) and UDP Glucuronosyltransferase 2 Family Polypeptide B15 (UGT2B15) are other transporters belonging to this increasing expression cluster with links to breast and prostate cancer respectively (Shehata *et al.*, 2008). Cysteine-rich Secretory Protein LCCL Domain Containing 1 (CRISPLD1, Section 4.3.1) is also a member of a family of proteins thought to have an association with the folate pathway (Chiquet *et al.*, 2011).

In the decreased expression cluster for the same time course (Cluster 5, Section 4.4.1.1) are a number of adhesion molecules, including *Cadherin-1* (*CDH1*), whose loss of function is thought to contribute to cancer progression by increasing proliferation, invasion, and metastasis, potentially through EMT (Ye *et al.*, 2012, Zhang *et al.*, 2013b) and *Epithelial cell adhesion molecule* (*EPCAM*), which is already in use as a target for immunotherapy treatment of human carcinomas (Martowicz *et al.*, 2012). More transport proteins are found in this cluster also, such as *CSAG2* (Section 4.3.1) whose expression has been shown to be predictive of clinical outcome in ovarian carcinoma patients (Materna *et al.*, 2007), suggesting that alterations in the influx and efflux of cisplatin in and out of the cells may be regulated in several ways.

The increasing and decreasing expression clusters (Clusters 6 and 3 respectively, Section 4.4.1.2) from the IC25 dose 1 week recovery time course using the same B-H adjusted *p*-value cut off of less than 1E-14, also yielded genes of potential interest. Of particular note is the presence of two genes, *EGFR* and *EML4*, which were previously discussed in Chapter 1 (Sections 1.6.1 and 1.6.3). Their highly significant differential expression in this time course indicates that well-known cancer pathways could be altered for survival in the presence of cisplatin.

Other genes associated with cancer include *CCND3* (Section 4.5.2), whose protein has been shown to interact with and to be involved in the phosphorylation of the recognised tumour suppressor protein, retinoblastoma protein (Rb). It is also thought the Cyclin-dependent kinase 4 (CDK4) activity associated with this particular cyclin protein is necessary for the cell's progression into the G2 phase of the cell cycle (Weng *et al.*, 2012), and so its reduced expression in this model may link with recent research that cisplatin treatment leads to an accumulation of cells in the G0/G1 phase (Barr *et al.*, 2013).

Top up-regulated differentially expressed genes in this time course have also been reported to be associated with various different types of cancer previously. *EGF-like repeats and discoidin 1-like domains 3* (*EDIL3*) is one such example with high expression levels predicting poor prognosis of patients in hepatocellular carcinoma (Sun *et al.*, 2010). In the cell model in this thesis it showed a two-fold change increase over the time course, while Secreted Frizzled-Related Protein 1 (SFRP1), a Wnt protein, may also delay the G1 phase and entry into the S phase through its downstream effectors on its increased expression (Ju *et al.*, 2009), and so could potentially be working to keep the cells in this phase of the cell cycle, similar to *CCND3*. Increased expression of other genes, Homeobox B2 (*HOXB2*, Table 4.10), Meis Homeobox 2 (*MEIS2*, Table 4.10), and Melanoma Antigen Family A 2 (*MAGEA2*), have also been associated with pancreatic cancer (Segara *et al.*, 2005), ovarian cancer (Crijns *et al.*, 2007), and progression in prostate cancer (Suyama *et al.*, 2010) respectively.

The 375 significant genes that overlap both treated time courses but are not significantly differentially expressed in the untreated (Section 4.5.2), could provide clinically useful information on the mechanisms of cellular survival in the presence of repeated cisplatin doses. Further network analyses will be performed on these genes to augment the GO analyses performed in Section 4.5.2. The presence of GO terms that include growth factor receptor pathways and response to stimuli suggests the treated cells are being pushed towards growth as a result of the treatment of the cells as well as the cells' own responses, so that understanding the system as a whole is vital.

The genes found in both IC25 and IC50 regimens are varied and include some well-known cancer genes such as *EGFR* and *JUNB* (Section 4.5.2); transporter proteins, e.g. transporter 1 ATP-binding cassette sub-family B (*TAP1*); adhesion molecules, e.g. basal cell adhesion molecule (*BCAM*), *L1CAM* (Section 4.5.2), melanoma cell adhesion molecule (*MCAM*, Table

4.10), and *NRCAM* (Sections 4.3.1 and 4.5.2); RAS proteins, e.g. RAB26 (Section 4.5.2), RAS guanyl releasing protein 3 (RASGRP3), ras homolog family member U (RHOU), and RAB38 (Section 4.5.2); and even the previously mentioned cell cycle protein CCND3 (Section 4.5.2), all of which are exciting findings. The combination of these genes may give an insight as to the key driving forces within the cell, many of which appear to indicate to the hallmarks of cancer discussed in Chapter 1 (Section 1.4). It is possible that the presence of transporter proteins in these lists may indicate they are playing a key role in helping the cells overcome the presence of cisplatin and survive its toxic effects. As discussed in Chapter 3 (Section 3.6), the cells may also be using other methods to escape cisplatin-induced death, such as the accumulation of cells in the G0/G1 phase of the cell cycle and the presence of potential effectors of that found here could support this theory (Barr *et al.*, 2013).

The overlap between treated and untreated time courses is reflected in their top similar functional annotation clusters generated by DAVID, i.e. that signalling molecules or glycoproteins are consistently the most enriched in the increasing or decreasing expression gene lists, independent of the time course they come from, whether treated or not, whether there are more time points, whether the dose is large or small, or whether the recovery is long or short. This shows that the key processes maintaining the cells through repeated passages of culture consistently rely on signalling, both within and between the cells in a population.

With the highly stringent threshold *p*-values, approximately 40% of genes overlap between the treated time courses and the untreated (Section 4.5.1) (41.5% IC50 dose 3 week recovery; 38.0% IC25 dose 1 week recovery) but it is noteworthy that few of these genes seem to share the same increasing or decreasing expression profiles over time (Table 4.7 and Table 4.8). This suggests that the treated and untreated cells may be under similar controls and regulators – the high overlap in differentially expressed genes – but could be acting in dissimilar ways

to produce different expression profiles for the same genes. This in turn results in the low numbers of overlapping genes within similar increasing or decreasing expression clusters over time.

Each of the 68 unique transcripts from Table 4.8 has been excluded from being defined as a gene of interest. A view of the whole system is however required for Network Analyses (see Chapter 5) and so they were retained in the list of differentially expressed genes used for these investigations.

## **5 Network Analysis**

### **5.1 Introduction to Network Analysis**

Global gene expression experiments capture an enormous amount of information with great accuracy. This information may be analysed in many different ways. In Chapter 4, the simplest form of analysis was used by identifying the genes that were the most significantly elevated or decreased at the end of the IC25 and IC50 dosing regimens. The use of clustering algorithms identified genes that shared distinctive patterns of up and down regulation over the entire course of the experiment. The membership of these clusters was determined by shared levels of expression at different time points. Genes may however be closely co-regulated but expressed at widely different levels.

Network analysis is an analytical tool that can be used to detect genes that are correlated in their expression, and will include genes that are negatively correlated as well as genes that have widely different absolute levels of expression. Summaries of the gene function (for example by GO analysis) of network members may reveal the function of the networks. Much can also be learned about co-ordinated gene function in a cellular system by examining network architecture, whereby highly connected genes (hubs) may identify suitable points for therapeutic intervention. The further level of information offered by network analyses could be of great value when investigating cisplatin resistance and has been explored here using weighted gene co-expression network analysis (WGCNA).

## 5.2 Weighted Gene Co-expression Network Analysis

Methods that consider the relationships and correlation patterns between transcripts have been successfully applied to a number of complex human traits (Russell *et al.*, 2003, Verdijk *et al.*, 2007). These methods, which elucidate the underlying transcriptional network structure, have allowed the identification of discrete modules of co-expressed, highly correlated, and potentially interconnected genes that associate with disease, intermediate disease-related phenotypes, external sample traits, or specific cell types (Smyth, 2004). Here the advanced network construction algorithm, Weighted Gene Co-expression Network Analysis (WGCNA), was performed using the same differential gene expression data generated in Chapter 4 (Section 4.3).

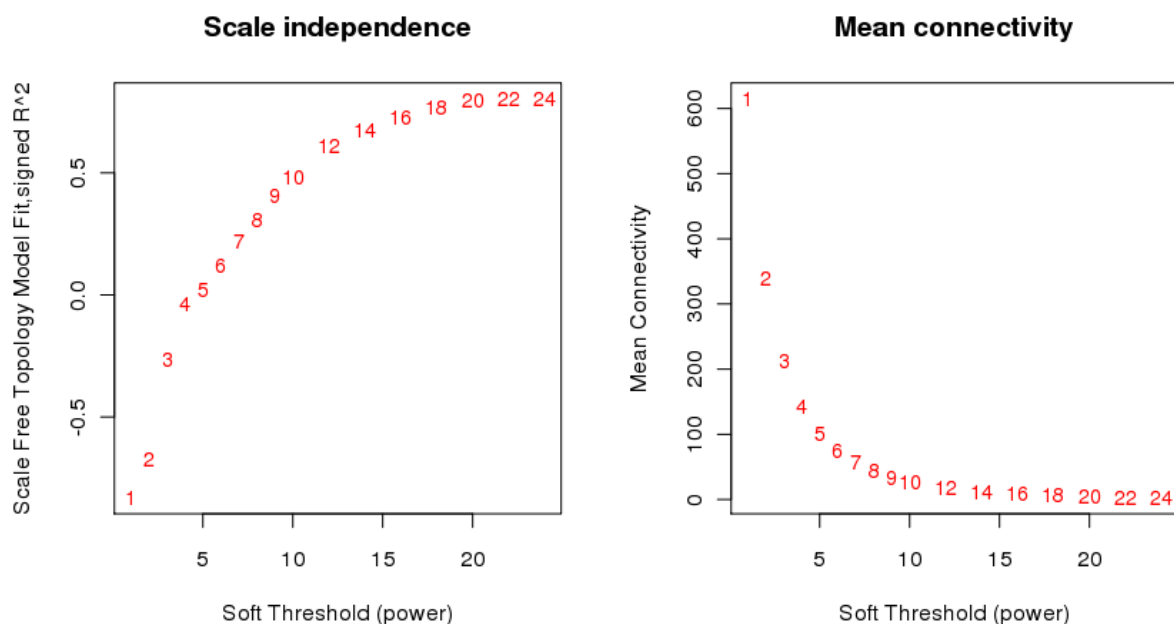
WGCNA involves several steps, the first of which is construction of a gene co-expression network that finds the interaction patterns among genes through their correlations. This correlation matrix is then used to identify modules using pathway-based analysis through hierarchical clustering (Section 5.2.1). The relationships between the trait of interest and the modules are compared (Section 5.2.2) and those of potential biological interest, with possible causal implications or particularly high connectivity can be investigated further. Here, the resultant modules have been related to inhibitory concentration value (IC<sub>25</sub> or IC<sub>50</sub> value) as a measure of cellular resistance at a particular time.

### 5.2.1 Creating the Network

For a weighted network, as is used in WGCNA, an adjacency matrix reports the connection strength between gene pairs. To generate the co-expression network, microarray gene expression data (pre-processed as outlined in Chapter 4 Section 4.2.3) can be used as the

basis of measuring the concordance of gene expression using Pearson correlations. The resultant matrix is then transformed continuously with a power adjacency function, which is approximately the equivalent of transforming data with highly skewed distributions before downstream analyses. The  $\beta$  statistic represents the power to which the correlation is raised.

The choice of  $\beta$  value requires the choice of a soft threshold power to which the co-expression similarity, i.e. the Pearson correlation, is raised to calculate adjacency. The lowest power for which the scale free topology fit index curve begins to flatten out after reaching a relatively high value was chosen (power 12 for IC50 dose 3 week recovery time course see Figure 5.1; power 10 for IC25 dose 1 week recovery time course). There is a trade-off between maximising scale free topology fit and maintaining a high mean number of connections; high signed  $r^2$  values may lead to networks with few connections. Mean connectivity is therefore also taken into account when choosing the value of  $\beta$ .



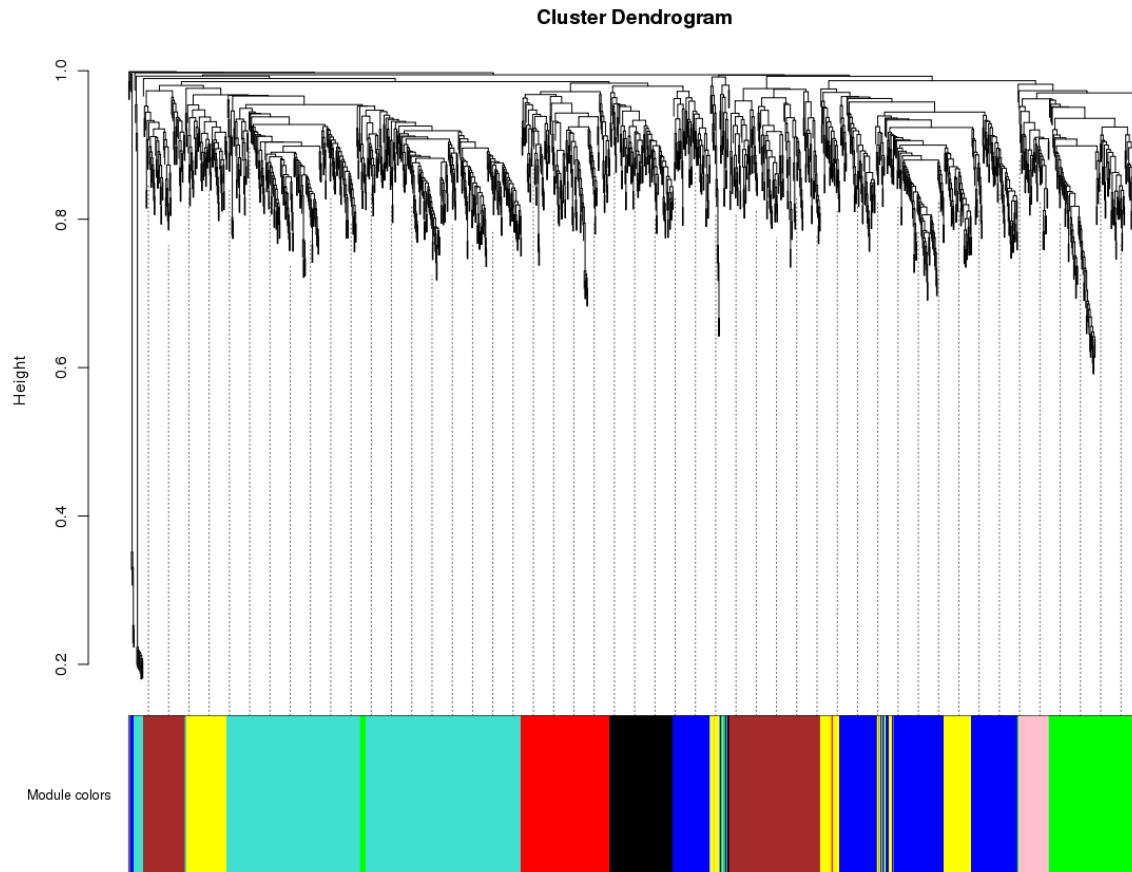
**Figure 5.1** A scale-free topology plot for the IC50 dose 3 week recovery time course. The power of 12 was used to perform the WGCNA algorithm used for the generation of modules. It can be seen that 12 also has a suitable mean connectivity (right hand plot) as the larger powers (14 and above) begin to have far fewer connections.

Once the adjacency matrix has been generated, network module definition can begin. Average linkage hierarchical clustering, coupled with a topological overlap dissimilarity measure has been used. The Topological Overlap Matrix (TOM) measures the overlap of the set of nearest neighbours of nodes or hubs, for example  $TOM_{ij}$  for hubs  $i$  and  $j$ . A value of 1 for  $TOM_{ij}$  would infer  $i$  and  $j$  are connected to the same substrates whereas a value of 0 would indicate that they do not share links to common substrates. The topological overlap dissimilarity ( $DistTOM_{ij}$ ) is used as input of hierarchical clustering (Langfelder and Horvath, 2008).

This combination of hierarchical clustering and topological overlap dissimilarity allows for hubs in the network that have a high topological overlap to be close to each other and for the identification of distinct modules. Potential hubs within a network are then associated with different modules, with varying strengths.

A dendrogram (Figure 5.2) was obtained by hierarchical clustering of the topological overlap dissimilarity with a height cut-off are used to arrive at clustering. The resultant modules correspond to the branches of the dendrogram and were assigned automatically on the datasets using Pearson's correlations.

Co-expression similarity of an entire module was calculated by their eigengene (the vector of the first principal component). These were used to identify transcript clusters which best match the module expression profile, summarised by its first principal component, and therefore may represent key genes within the pathway. All modules with genes that were highly co-expressed or contained fewer than 30 genes were merged. Eigengenes are also used as the vector to seek relationships between modules and a range of clinically relevant physiological and biochemical parameters.



**Figure 5.2** The hierarchical cluster dendrogram showing the assignment of genes to modules (colour band) based on the TOM for the IC50 dose 3 week recovery time course.

### 5.2.2 Relating Modules with Traits of Interest

In order to identify modules that are significantly associated with a particular trait, for example IC value as a measure of resistance, the eigengenes generated can be correlated quantitatively with the external trait of interest (Figure 5.3). Associations of individual genes with the trait of interest can then be quantified by defining Gene Significance (GS) as the absolute value of correlation between a particular gene and the trait. The larger the GS value is, the more biologically significant it is for that gene and this in turn gives rise to a module significance measure.

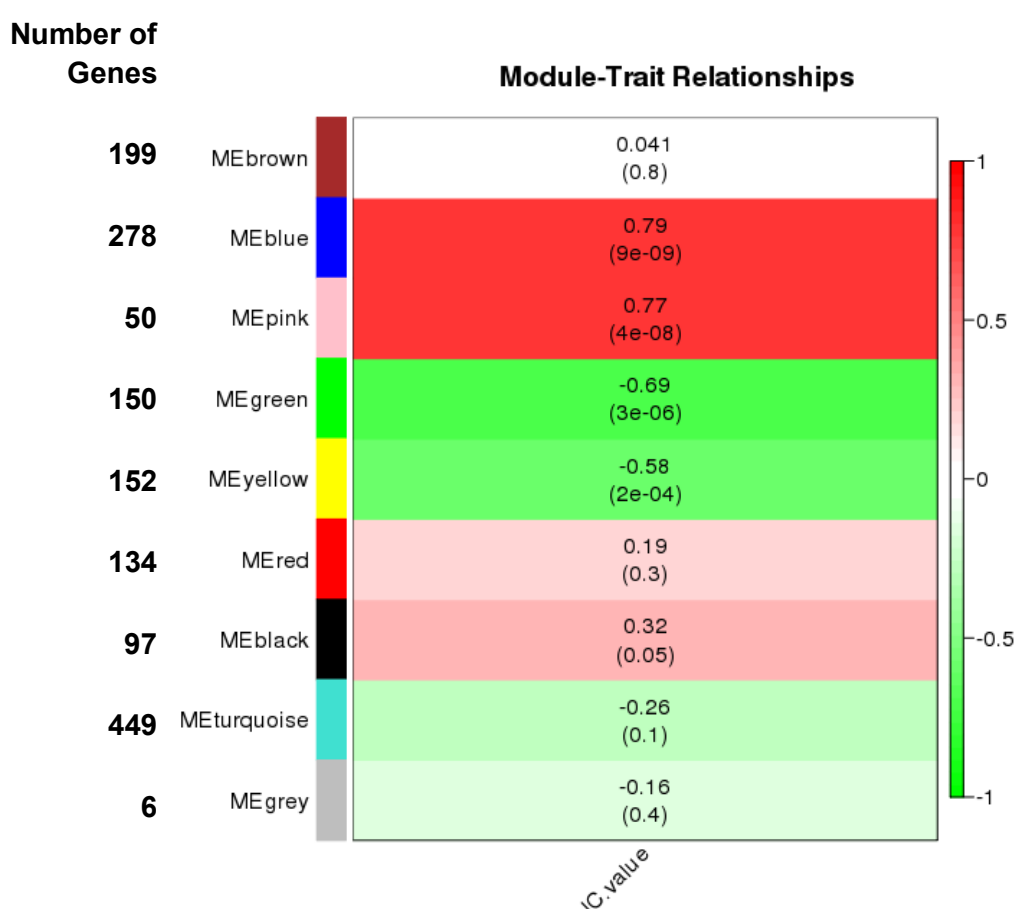
A quantitative measure of module membership (MM) as the correlation of the expression profile of the gene and the module eigengene can also be derived. This enables the quantification of the similarity of all genes to each module. This statistic,  $k_i$ , is closely related to the concept of intra-modular connectivity (kME) and is a useful indicator of hub gene status that may be used to detect key driver genes in modules of interest for a particular trait. MM is a measure of kME and so the higher it is, the more likely the gene is to be a key driver in that module.

The most significantly differentially expressed genes within a certain module and their values within the TOM are selected out and used to generate Circos plots (Section 5.2.3) to visualise the connections. Circos plots are used to map this connectivity in a quantified way; red connections show high correlation (98-100%), decreasing the correlations through the spectrum (dark orange >96%, yellow >94%, green >92%, royal blue >90%) through to pale blue connections, showing a lower correlation (80-90%). For interpretation ease, where the number of genes within a module were 50 or less, the Circos plots were created using the entire module but for modules with larger numbers of genes, the most significant differentially expressed genes ( $p$ -value less than  $1E-14$ , same threshold as applied in Chapter 4 [Section 4.4]) have been used. Combining this representation of connectivity with the top kME genes allowed the top hubs within each module of interest to be identified.

The emphasis on modules and pathways instead of individual genes in WGCNA greatly alleviates the problem of multiple comparisons. The use of gene expression data for module definition ensures no prior pathway information is used and the use of the power of correlations allows for a unified approach for relating variables of interest.

### 5.2.3 IC50 Dose 3 Week Recovery Time Course Modules of Interest

The 1,515 genes, found to be differentially expressed in the IC50 dose 3 week time course, (Chapter 4 Section 4.3) were subject to WGCNA. A total of eight co-expression modules were identified using the automatic network construction algorithm and each module was assigned a unique identifier (the name of a colour). A total of six transcript clusters remained unassigned, showing no significant correlation with any of the observed modules (Figure 5.3, full gene lists E-Appendix 9.2.1).



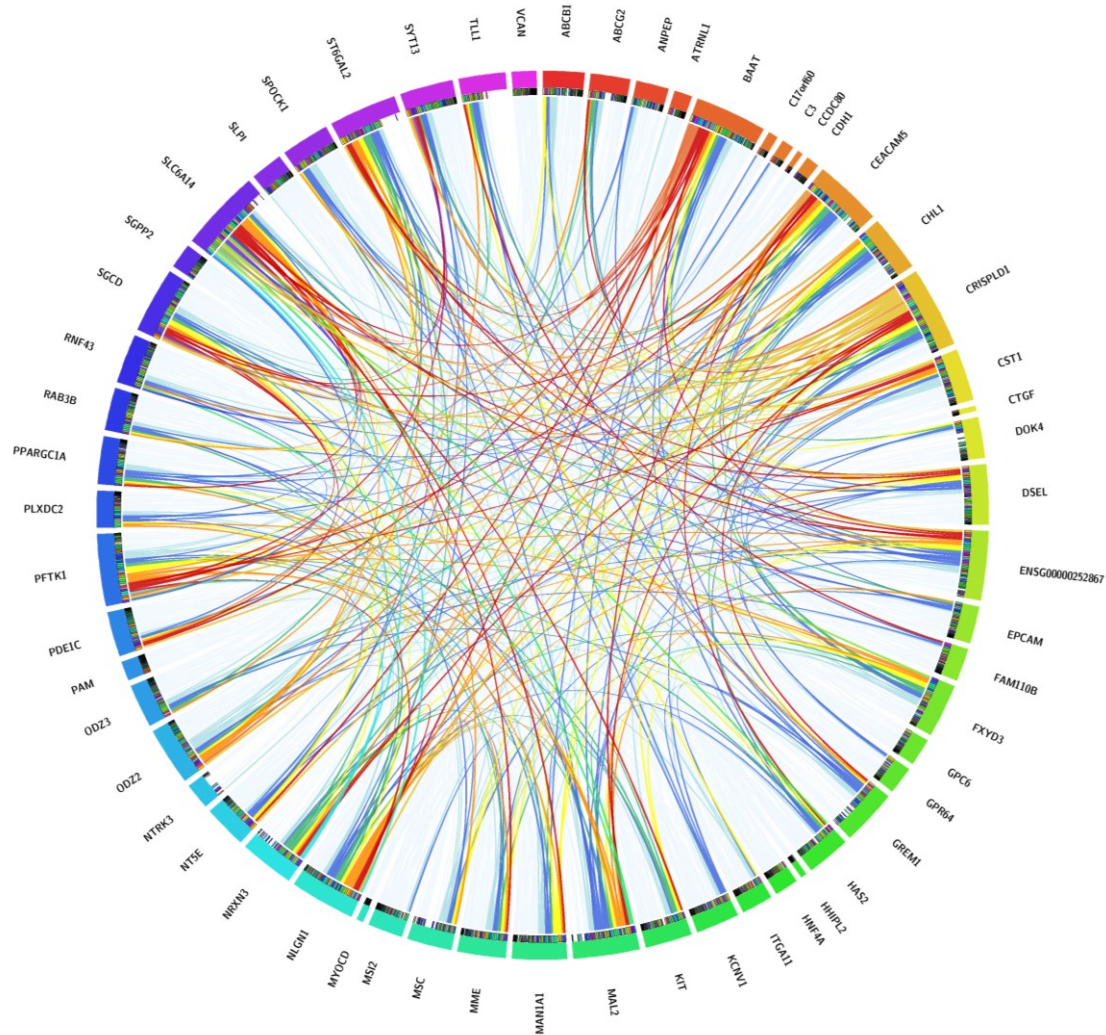
**Figure 5.3** A module-trait relationship matrix for the IC50 dose 3 week recovery time course modules. The intensity of the red and green colour corresponds to positive and negative correlation respectively of that module with the IC50 value. Of particular interest here are the blue and pink modules that show significant positive correlation whereas the green and yellow modules show significant negative correlation with IC50 value. The number of genes within each module is shown alongside.

When examining Figure 5.3, it can be seen that the blue and pink modules had the highest positive correlation with the IC<sub>50</sub> value ( $r^2$  0.79  $p$ -value 9E-09;  $r^2$  0.77  $p$ -value 4E-08 respectively) while the green and yellow modules had the highest negative correlation ( $r^2$  -0.69  $p$ -value 3E-05;  $r^2$  -0.58  $p$ -value 2E-04 respectively). Further analyses conducted were therefore focused around these four key modules.

GO analysis of the genes contained within each module showed strong enrichment for signalling peptides and glycoproteins for both the blue and green modules. The blue module also showed significant levels of “EGF-like domains” (B-H adjusted  $p$ -value 2.08E-04) and localisation to the plasma membrane region (GO:0031226 B-H adjusted  $p$ -value 0.0015). The green module genes appeared to have preferential locations to the plasma membrane region and the extracellular space although these did not reach significance (GO:0005886 B-H adjusted  $p$ -value 0.039; GO:0005576 B-H adjusted  $p$ -value 0.017 respectively). The yellow module showed enrichment for lipid biosynthesis and membrane fraction location (GO:0008610 B-H adjusted  $p$ -value 0.014; GO:0005624 B-H adjusted  $p$ -value 0.055 respectively). The pink module, however, produced no significant functional annotation clusters.

Circos plots of connectivity for the most significant genes within each module were generated to visualise potential hubs and their interactions within the module. Figure 5.4 is the Circos plot for the blue module (see Appendix 9.3.1 for those of other modules from the IC<sub>50</sub> Dose 1 Week Recovery Time Course). Certain genes appeared more highly connected than others, both in number of connections, indicated by the number of lines emanating from each gene’s segment and strength of correlation. For a connection to be seen on the Circos plot the correlation has to be over 80% and the strength of this correlation is indicated by the colour of the connection; red for the strongest (greater than 98%), pale blue for the weakest

(between 80%-90%). Some genes appear to be highly connected with several strong connections, such as BAAT, CRISPLD1, the non-coding RNA ENSG00000252867 (Chromosome 5), MAL2, PFTK1, SGCD, SLC6A14, and ST6GAL2. Interestingly two of the most highly connected genes: CRISPLD1 and epithelial marker/transmembrane protein MAL2 (Figure 5.4) were found in the increasing temporal expression cluster (Cluster 2) over the course of cisplatin treatment (Chapter 4 Sections 4.3.1 and 4.6).



**Figure 5.4** Circos plot of the Blue module from the IC50 dose 3 week recovery time course. Several strong correlations are visible between the 57 genes on the plot, as shown by the colour intensity. Several transcripts have multiple connections (several lines) to other transcripts, some of which are especially high in correlation as shown by their red or orange colour.

In order to verify mathematically what can be seen visually on the Circos plots,  $kME_{\text{module}}$  values (a measure of eigengene-based connectivity) were used to find potential hubs, mathematically (Table 5.1). As expected, the visual connection hubs seen in the Circos plots (Figure 5.4 and Appendix 9.3.1) are present in the genes with the greatest  $kME$  values, such as *ENSG00000252867*, *SLC6A14*, and *MAL2* in the blue module; *SLC40A1* and *SCN9A* in the pink module; *LPPR1* and *CDH19* in the green module; and *CACNG6* and *TP53INP1* in the yellow module.

| Module | High Connectivity Genes<br>( $kME$ for module >0.92)                                                                                                                                                                                                                                                                                             |
|--------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Blue   | <i>ENSG00000252867</i> , <i>SLC6A14</i> , <b><i>FOXQ1</i></b> , <i>CRISPLD1</i> , <i>BAAT</i> , <i>GREM1</i> , <i>MAL2</i> , <i>ST6GAL2</i> , <i>CHL1</i> , <i>PFTK1</i> , <i>DSEL</i> , <i>MAN1A1</i> , <i>ENPP4</i> , <i>AFAP1</i> , <i>DIRAS3</i> , <i>SLC27A6</i> , <i>JAM3</i> , <i>MSX2</i> , <i>SLC16A9</i> , <i>MMP16</i> , <i>MYO5B</i> |
| Pink   | <i>CDCP1</i> , <i>SLC40A1</i> , <i>SCN9A</i> , <i>CXCL1</i> , <i>PSMB9</i> ,                                                                                                                                                                                                                                                                     |
| Green  | <i>LPPR1</i> , <i>CDH19</i> , <i>ACSS3</i> , <i>ITM2A</i> , <i>VSIG1</i> , <i>ANKS4B</i> , <i>UGT2B15</i>                                                                                                                                                                                                                                        |
| Yellow | <i>CADPS2</i> , <i>CACNG6</i> , <i>SCEL</i> , <i>SPARC</i> , <i>CDH6</i> , <i>TP53INP1</i> , <i>TRIM29</i> , <i>BMP7</i> , <i>ZNF717</i> , <i>PCDHB5</i> , <i>NFKBIZ</i> , <i>ST3GAL5</i> , <i>CARD8</i> , <i>ETHE1</i> , <i>FDFT1</i> , <i>DHTKD1</i> , <i>TP53</i>                                                                             |

**Table 5.1** Most highly connected genes within modules that have the most significant associations with IC50 value. The genes in bold correlate with particularly high  $kME$  values of greater than 0.95 (0.97 for the green module).

Across the four modules of interest generated in the WGCNA, there are genes that were also found in the increasing and decreasing temporal expression clusters (Chapter 4 Section 4.4, E-Appendices 9.1.3 and 9.1.4; Table 5.2; E-Appendix 9.2.2). The change in expression over the time course of any of these hub genes could be of great interest, not just for their individual effects but also for the way they affect other substrates in their pathways. Genes of particular interest from these investigations will be discussed later (Section 5.3).

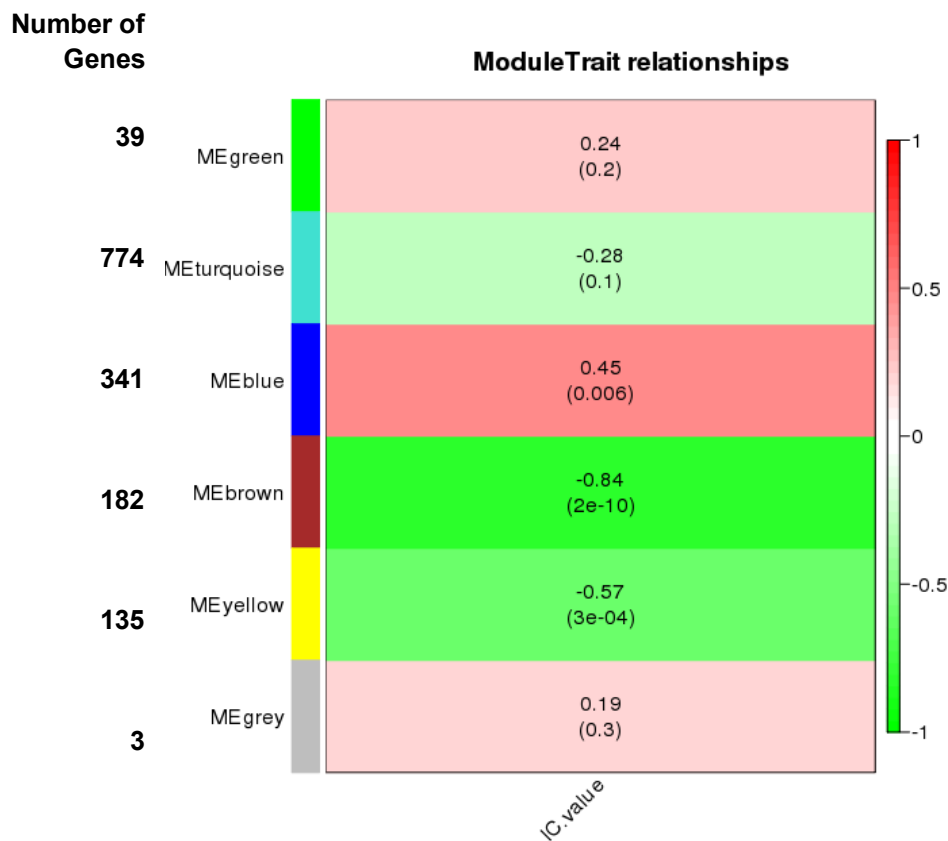
| <b>IC50 Dose<br/>3 Week Recovery</b> | <b>Blue (278)</b> | <b>Pink (50)</b> | <b>Green (150)</b> | <b>Yellow (152)</b> |
|--------------------------------------|-------------------|------------------|--------------------|---------------------|
| <b>Up (227)</b>                      | 42                | 6                | 24                 | 14                  |
| <b>Down (209)</b>                    | 36                | 10               | 14                 | 22                  |

**Table 5.2** Number of genes showing increasing (up) and decreasing (down) expression from the temporal expression clustering, conducted in Chapter 4 Section 4.4, across the modules of greatest correlation with IC50 value from the IC50 dose 3 week recovery time course.

#### 5.2.4 IC25 Dose 1 Week Recovery Time Course Modules of Interest

The 1,474 genes found to be differentially expressed in the IC25 dose 1 week time course (Chapter 4 Section 4.3) were subject to WGCNA. A total of five co-expression modules were identified using the same algorithm for automatic network and module generation (Section 5.2.3), using a power of 10 for the generation of the adjacency matrix via the use of a scale free topology plot. Each module was again identified by a unique colour and a total of three transcript clusters remain unassigned, showing no significant correlation with any of the observed modules.

Only three of the modules were found to be significantly correlated with the IC25 value (Figure 5.5, for module full gene lists see E-Appendix 9.2.3). The blue module was positively correlated ( $r^2$  0.45  $p$ -value 0.006) whilst the brown and yellow modules were negatively correlated ( $r^2$  -0.84  $p$ -value 2E-10;  $r^2$  -0.57  $p$ -value 3E-04 respectively).



**Figure 5.5** A module-trait relationship matrix for the IC25 dose 1 week recovery time course modules. The intensity of the red and green colour corresponds to positive and negative correlation respectively of that module with the IC25 value. Of particular interest here is the blue module that shows significant positive correlation whereas the brown and yellow modules show significant negative correlation with IC25 value. The number of genes within each module is shown alongside.

GO analysis of the genes contained within each module showed strong enrichment for signalling peptides and glycoproteins for both the brown and yellow modules (B-H adjusted *p*-value 0.0027 and B-H adjusted *p*-value 6.57E-04 respectively). The blue module, although not significant, showed partial enrichment for cell adhesion processes (GO:0007155).

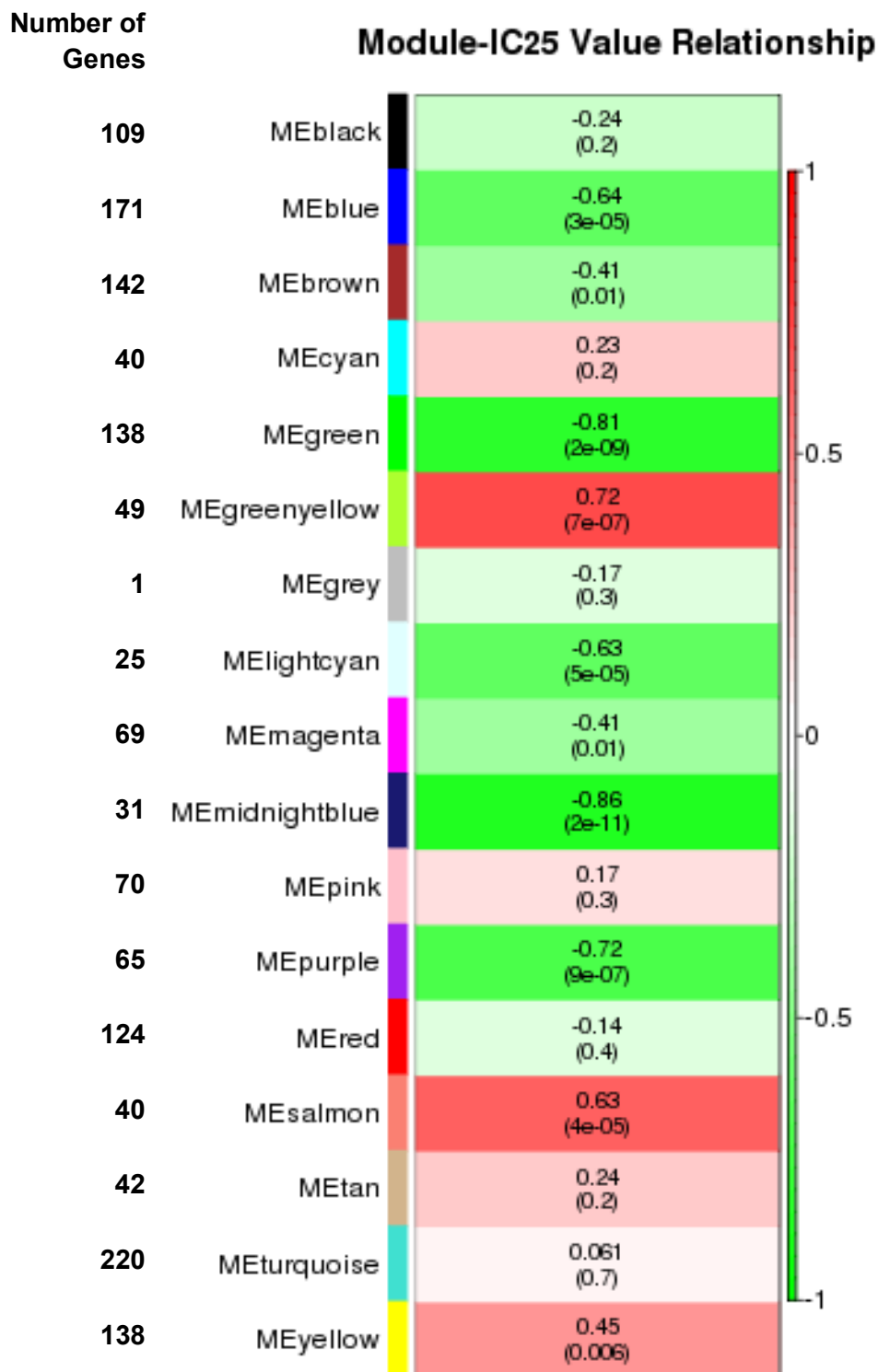
Circos plots were generated for these modules of interest to visualise potential hub genes of interest (Appendix 9.3.2) and there were genes with several, highly correlated connections that are potential hubs. Interestingly within the blue module, these potential hubs include well-known genes associated with treatment implications in lung adenocarcinomas, for example EGFR and EML4 (Chapter 1 Sections 1.6.1 and 1.6.3).

As seen for the IC50 dose 3 week time course (Section 5.2.3), there appeared to be several significantly differentially expressed genes that lie within either increasing or decreasing temporal clusters (Chapter 4 Section 4.4; E-Appendices 9.1.3 and 9.1.4), in each of the three modules of interest in the IC25 dose 1 week recovery time course (Table 5.3, E-Appendix 9.2.4). The difference in proportion between the two time courses may be due to the shorter time between cisplatin exposure and extraction of RNA or due to the fewer number of modules generated from the automatic WGCNA algorithm. Genes of particular interest from these investigations will be discussed later (Section 5.3).

| <b>IC25 Dose<br/>1 Week Recovery</b> | <b>Blue (349)</b> | <b>Brown (319)</b> | <b>Yellow (171)</b> |
|--------------------------------------|-------------------|--------------------|---------------------|
| <b>Up (285)</b>                      | 72                | 46                 | 30                  |
| <b>Down (323)</b>                    | 73                | 73                 | 31                  |

**Table 5.3** Number of genes showing increasing (up) and decreasing (down) expression from the temporal expression clustering, conducted in Chapter 4 Section 4.4, across the modules of greatest correlation with IC25 value generated by the automatic WGCNA module generation algorithm from the IC25 dose 1 week recovery time course.

Due to the large number of genes within modules, e.g. 774 within the turquoise module (Figure 5.5), a more stringent level of hierarchical clustering was applied to these modules. Conducting the same procedure on the data from the IC50 dose 3 week time course led to no changes in module generation. Seventeen modules were produced and correlated with the IC25 value. It can be seen from Figure 5.6 that the modules with greatest gene significance for IC25 value are the midnight blue, green, green yellow and purple modules (for full gene lists of all modules, see E-Appendix 9.2.5). The green yellow module has the highest positive correlation ( $r^2$  0.72  $p$ -value 7E-07) while the midnight blue, green, and purple have the highest negative correlations ( $r^2$  -0.86  $p$ -value 2E-11;  $r^2$  -0.81  $p$ -value 2E-09;  $r^2$  -0.72  $p$ -value 9E-07 respectively). These modules will be the focus in the search for potential hub genes.



**Figure 5.6** A module-trait relationship matrix for the IC25 dose 1 week recovery time course modules generated by cutting higher on the clustering dendrogram, reducing the number of modules. The intensity of the red and green colour corresponds to positive and negative correlation respectively of that module with the IC25 value. Of particular interest are the midnight blue, green, and purple modules that show significant negative correlation whereas green yellow module shows significant positive correlation with IC25 value. The number of genes within each module is shown alongside.

As outlined in Section 5.2.2,  $kME_{\text{module}}$  values were used to mathematically find potential hubs within these four modules (Table 5.4). As is to be expected, the high connectivity genes in these modules overlap heavily with those from the automatic network construction modules.

| Module        | High Connectivity Genes<br>( $kME$ for module $>0.95$ )                                                  |
|---------------|----------------------------------------------------------------------------------------------------------|
| Midnight Blue | <i>CCDC80, TP63, CA2, NEDD9, RH3HDML, SULT2B1, LSR, FOXA3, TGFBRI, UBASH3B, ZNF850P, CNTNAP1</i>         |
| Green         | <i>BAAT, GLI3, TGM2, DMKN, ITGA11, SDR42E1,</i>                                                          |
| Green Yellow  | <i>ETS2, ZNF616, DOK6, PDLIM7, EML4, LPCAT2, UTY, KDM5D, DENND5B</i>                                     |
| Purple        | <i>DPYSL3, C21orf63, CA8, A1CF, ITM2A, DCDC5, RH3HDML, CADM2, PPP1R1C, CRABP1, HHLA2, HNF4A, DEDDD2D</i> |

**Table 5.4** Most highly connected genes within the reduced modules that have the most significant associations with IC25 value.

The decrease in the number of genes within the modules of interest subsequently leads to a decrease in the proportion that appear in the increasing and decreasing temporal expression clusters over the course of cisplatin treatment (Chapter 4 Section 4.4, Table 5.5, E-Appendix 9.2.6) compared with earlier in this section (Table 5.3). Using these smaller modules allows more specific potential genes of interest to be elucidated. Genes of particular interest from these investigations will be discussed later (Section 5.3).

| IC25 Dose<br>1 Week Recovery | Midnight Blue<br>(31) | Green (138) | Green Yellow<br>(49) | Purple (65) |
|------------------------------|-----------------------|-------------|----------------------|-------------|
| Up (285)                     | 6                     | 23          | 9                    | 8           |
| Down (323)                   | 8                     | 31          | 12                   | 12          |

**Table 5.5** Number of genes showing increasing (up) and decreasing (down) expression from the temporal expression clustering, conducted in Chapter 4 Section 4.4, across the modules of greatest correlation with IC25 value for reduced WGCNA modules from the IC25 dose 1 week recovery time course.

### 5.2.5 Normal Untreated A549 Cells Time Course

In order to perform comparisons, the 865 genes found to be differentially expressed in the normal untreated time course (Chapter 4 Section 4.5.1) was subject to WGCNA. Three co-expression modules were identified (each assigned a unique colour) with ten transcript clusters remain unassigned (Table 5.6, E-Appendix 9.2.7).

| Module      | Blue | Brown | Grey | Turquoise |
|-------------|------|-------|------|-----------|
| No of Genes | 278  | 154   | 10   | 423       |

**Table 5.6** The number of genes within each module colour after the WGCNA algorithm has been performed on the 865 differentially expressed genes from the normal time course. The grey module contains genes that cannot be assigned to a module. Full gene lists are available in E-Appendix 9.2.7.

GO analysis was performed to see what processes were enriched in the normal time course modules. All three modules involve a membrane-associated GO term (Table 5.7) and a glycoprotein (turquoise and blue modules) or signal (brown module) term, suggesting that these are the main processes within normal A549 cells not under the selection pressure of the presence of cisplatin. The presence of these GO terms in the untreated time course WGCNA modules suggests that it may be worth focusing on the functional annotation clusters that do not match these exactly in the treated time courses WGCNA (Sections 5.2.3 and 5.2.4). The lack of references to growth receptors, regulation of differentiation, responses to various stimuli, and similar terms that have been seen in previous GO analyses (Chapter 4 Sections 4.4.1.1, 4.4.1.2 and 4.5.3; Chapter 5 Sections 5.2.3 and 5.2.4) suggests these may be the key terms that result from cisplatin exposure.

| Module    | GO Term                                | Enrichment | Gene Count | B-H <i>p</i> -value |
|-----------|----------------------------------------|------------|------------|---------------------|
| Turquoise | Plasma membrane                        | 9.4        | 100        | 2.9E-10             |
|           | Glycoprotein                           | 8.13       | 160        | 2.1E-12             |
|           | Plasma membrane                        | 5.76       | 145        | 1.2E-10             |
|           | Cell migration, motion, and locomotion | 5.2        | 18         | 7.0E-04             |
| Blue      | Glycoprotein                           | 5.53       | 96         | 6.3E-05             |
|           | Membrane fraction                      | 3.82       | 30         | 2.8E-03             |
|           | Complement and coagulation cascade     | 2.27       | 10         | 3.1E-04             |
| Brown     | Plasma membrane                        | 4.51       | 37         | 2.7E-03             |
|           | Signal                                 | 3.83       | 49         | 1.6E-04             |
|           | Cell adhesion                          | 1.99       | 11         | 4.4E-02             |

**Table 5.7** Highly enriched GO term functional annotation clusters for the modules created by the automatic network generation WGCNA algorithm for the 865 differentially expressed genes from the normal untreated time course.

#### 5.2.6 Differentially Expressed Genes in both Treated Time Courses but not present in the Untreated Time Course

Previous analyses (Chapter 4 Section 4.5.3) that had taken into consideration changes in expression observed without cisplatin treatment found that 375 genes were differentially expressed in both treatment time courses when taking the normal time course into account. It is conceivable that key genes involved in evasion of cisplatin-induced cell death are contained within this set of genes and that their interactions together and with other genes may be highly biologically relevant.

The RMA data from the IC50 dose 3 week recovery time course was used as the expression data required for the WGCNA, rather than the IC25 dose 1 week recovery time course, because of the smoother increase in IC50 value over the course of cisplatin exposure (Chapter 3 Section 3.4) and for the matching of time points with the normal untreated time course. The adjacency matrix was created using a power of 26, according to scale free topology (Section 5.2.1) and the automatic network generation algorithm was used to give the modules outlined in Figure 5.7. Fifteen transcript clusters remained unassigned, showing no significant correlation with any of the observed modules.

As previously described, the eigengenes of these modules were correlated with increasing IC50 value. From Figure 5.7, it can be seen that the green, yellow, and blue modules have the highest positive correlation with the IC50 value ( $r^2$  0.82  $p$ -value 11E-09;  $r^2$  0.75  $p$ -value 1E-07;  $r^2$  0.59  $p$ -value 1E-04 respectively). It is noteworthy that all modules with a significant association with IC50 value have a positive correlation.



**Figure 5.7** A module-trait relationship matrix for the 375 genes only differentially expressed in the treated time courses using the IC50 dose 3 week recovery time course data. The intensity of the red and green colour corresponds to positive and negative correlation respectively of that module with the IC50 value. Of particular interest here are the green, yellow and blue modules that show significant positive correlation with IC50 value. The number of genes within each module is given alongside.

MM within the yellow module was found to exhibit a positive correlation with the gene significance for increasing IC50 value (cor= 0.66,  $p$ -value = 6E-07), confirming this may be meaningful biologically and suggest the presence of hub genes. The same can be said for the MM of the blue module (cor= 0.45,  $p$ -value = 1.7E-04) and the blue module (cor= 0.62,  $p$ -value = 1.5E-05).

In order to investigate the potential presence of hub genes within the three modules of interest, Circos plots (Appendix 9.3.3) and measures of intra-modular connectivity, kME (Table 5.8), were generated. These modules show a much higher level of connectivity and

strength of correlation between the module members than has been seen previously. GO analyses however yielded no significant functional annotation clusters.

| <b>Module</b> | <b>High Connectivity Genes<br/>(kME for module &gt;0.9)</b>                                              |
|---------------|----------------------------------------------------------------------------------------------------------|
| <b>Green</b>  | <i>TLL1, DSEL, SGCD, FGFR1, PCDH7, CCDC146, JAM3, NID2</i>                                               |
| <b>Yellow</b> | <i>CDCP1, DCDC5, SCN9A, BTBD11, L1CAM, RHOU, FAS, TAP1, PLCD3</i>                                        |
| <b>Blue</b>   | <i>HHLA2, SLN, C21orf63, ANKRD30A, TBC1D2, PRKCH, OR51B4, CTSH, NRCAM, LPAR5, R3HDML, SPATA18, PTPRG</i> |

**Table 5.8** Most highly connected genes from the 375 overlapping genes within modules that have the most significant associations with IC50 value.

### 5.3 WGCNA Discussion

The focus of this chapter has been to explore further the differential gene expression findings for the two cisplatin time courses by using WGCNA to identify modules of genes that may be under co-regulation or working together. There are a number of genes, identified through these network analyses which are of particular interest due to prior studies implicating them in cancer pathways and phenotypes, and drug transport.

For both time courses, genes were found to overlap both modules with an association for the IC value and the increasing and decreasing temporal expression clusters (Chapter 4 Section 4.4). Any potential hubs that show increasing or decreasing expression over cisplatin rounds could be important genes. As discussed earlier in this chapter (Sections 5.2.3 and 5.2.4), combining information gleaned from temporal expression clusters and the modules generated from WGCNA on genes of interest, it is possible to begin to build a picture of those genes that may be exerting a driving force for survival of cells when exposed to cisplatin in culture.

Eight modules were generated using the data from IC50 dose 3 week recovery time course and those of interest that show correlation with the IC50 value (blue, pink, green, and yellow) all contain genes that have previously been implicated in various cancers. These genes are also within temporal expression clusters of interest (Chapter 4 Section 4.4.1.1)

The blue module from the IC50 dose 3 week recovery time course has a large proportion of the genes that increase and decrease in correlation with cisplatin treatment and there were a couple of genes that could be of particular interest that were highly significantly differentially expressed (B-H adjusted  $p$ -value less than  $1E-14$ ). *ABCG2* that encodes the Breast Cancer Resistance Protein (BCRP), known to be associated with drug transport and efflux (Sukowati *et al.*, 2012), showed a significant increase in expression over the time course, while *GPC6* decreased over the course of treatment. Interestingly a homolog of GPC6, GPC1 has been implicated in pancreatic cancer (Whipple *et al.*, 2008).

In the pink module, the Tenascin C gene was seen to increase over the course of cisplatin treatment and previously its product has been shown to enhance cancer cell growth and motility, affecting cell adhesion through integrin pathway activation (Paron *et al.*, 2011).

The green module contained many of the genes with the smallest B-H adjusted  $p$ -values with increases and decreases in expression over the cisplatin treatment time. Some have already been shown to have effects on tumour progression and invasiveness on increased expression including UGT2B15 that has been shown to have differential protein expression in prostate tumour cells during cancer progression (Pâquet *et al.*, 2012). The taxol-resistance-associated gene *CSAG2* also had significant differential expression within this module and is known to be predictive for clinical outcome in ovarian carcinoma patients (Materna *et al.*, 2007).

In the yellow module, a further three genes of potential interest were found to have either highly significant differential increases or decreases in expression. *VSNLI*, also known as *VILIP1*, is a putative suppressor that counteracts EGF-induced epithelial-mesenchymal transition (EMT) in squamous carcinoma cells (Schönrath *et al.*, 2012) and is known to be down-regulated during NSCLC mechanisms (Fu *et al.*, 2008), as has been seen in the time course experiment of this present study. The other two genes are: *TP53INP1* a known target for a miRNA, miR-17-5p, that regulates cell proliferation and apoptosis in cervical cancer (Wei *et al.*, 2012) cells and *BMP7* expression has been proposed to possibly affect lung cancer cell invasion ability, potentially with clinical significance (Liu *et al.*, 2012).

The same is true for the IC25 dose 1 week recovery time course. Five modules were generated using the automatic network construction and those of interest showing correlation with the IC25 value (blue, brown, and yellow) all contain genes that have previously been implicated in various cancers. Several of these are also within temporal expression clusters of interest (Chapter 4 Section 4.4.1.2).

The blue module contained many genes known to be associated with cancer pathways, such as *EML4* (Chapter 1 Section 1.6.3), *TNS4* a target of KRAS signalling (Al-Ghamdi *et al.*, 2011), and *SGK1* that is known to be required for mTORC1 activity in breast cancer (Hall *et al.*, 2012), as well as decreasing drug sensitivity of colon cancer cells (Amato *et al.*, 2012). Another potentially important gene in this module is *PDXK*. *PDXK* encodes the protein that phosphorylates vitamin B6, a step key in its metabolism, which may influence intracellular accumulation of cisplatin (Galluzzi *et al.*, 2013a) and may have effects on oncogenesis, tumour progression and therapeutic responses (Galluzzi *et al.*, 2013b). *SLIT3* (Dickinson *et al.*, 2011), *MEIS2* (Crijns *et al.*, 2007) and *FOSL2* (Hein *et al.*, 2009) also have altered

expression in this module over the time course and have been associated with ovarian cancers and malignant potential.

Interestingly several cancer-related-pathway genes were found to have significant differential expression in the brown module. SPARC protein expression has already been shown to correlate with clinical characteristics of NSCLC (Huang *et al.*, 2012) and IGFBP7 down-regulation has been observed in cisplatin-resistant NSCLC (Okamura *et al.*, 2012) so the decreasing expression of these two genes over the time course, is not wholly unexpected. Other genes with previous reports of association with NSCLC were also present in this module with significant changes in expression. These include *CDH12*, associated with NSCLC promotion and progression (Bankovic *et al.*, 2010), potentially by enabling invasion and *HMOX1*, associated with NSCLC tumour invasiveness and poor clinical outcome (Tsai *et al.*, 2012), potentially through autophagy- and apoptosis-related molecules (Banerjee *et al.*, 2012). Similar to findings for the other modules, genes associated with other cancers as novel biomarkers were also observed in the brown module such as *DMKN* a prognostic indicator of early stage colorectal cancer (Tagi *et al.*, 2010). Genes and their products with links to ovarian cancer also appear: *PAQR5* (Charles *et al.*, 2010), *TUSC3* (Pils *et al.*, 2013), *FILIP1* (Xie *et al.*, 2011), and *SCN9A* (Yildirim *et al.*, 2012).

In the yellow module, again there were several genes with significant differential expression of potential interest. *ALX1* is known to promote EMT through *SNAIL* expression and has been found in global expression changes for A549 cells exposed to anti-cancer drugs (Yuan *et al.*, 2013). Constitutive activity of other members of this module, found to have increasing expression over the cisplatin time course, have also been associated with poor prognosis in other cancers: *EDIL3* (Sun *et al.*, 2010), *LPAR1* (Marshall *et al.*, 2012), *TNC* (Paron *et al.*, 2011), *SFRP1* (Elzi *et al.*, 2012). Hypo-expression of other members *CADM2* (Chang *et al.*,

2010) and *PLA2G4A* (Zhang *et al.*, 2013a) are thought to correlate with unfavourable outcomes in prostate and gastric cancers respectively.

Modules generated with the greater hierarchical clustering stringency also contain genes of biological interest. The midnight blue module, , contains a few genes that have already been highlighted as potential markers for cancer prognosis (MAP7 [Blum *et al.*, 2008], TSPAN7 [Wuttig *et al.*, 2012], and CXXC5 also known as RINF [Knappskog *et al.*, 2011]). SIK1 is also present in this module. It is induced by TGF beta to negatively regulate type I receptor kinase signalling (Kowanetz *et al.*, 2008) and couples the known cancer gene LKB1 (Chapter 1 Section 1.6.7) to induce p53-dependent cell-detachment apoptosis (Cheng *et al.*, 2009), and acts as a tumour suppressor and prevents metastasis (Shaw, 2009).

Genes in the green module have already been associated with various cancers and even responses to treatment. Two up-regulated genes have been identified as potential markers in chemo-resistant epithelial ovarian cancer (RNF125 [Ju *et al.*, 2009]) and for NSCLC (KRT15 [Sanchez-Palencia *et al.*, 2011]). ERGIC1 has also been identified as a potential drug target in prostate cancer (Vainio *et al.*, 2012). Potential oncogenes and tumour suppressors are also found in this module. RASGRP3 has been found to contribute to the tumorigenic phenotype in human melanoma (Yang *et al.*, 2011) as well as forming and maintaining of the prostate cancer phenotype (Yang *et al.*, 2010) and a knockdown of the FGFBP1 has been shown to have anti-tumour effects on knockdown in colon carcinoma (Schulze *et al.*, 2011). RASSF8 (Lock *et al.*, 2010), ZFP36 (Upadhyay *et al.*, 2013), and CRIP2 (Cheung *et al.*, 2011) appear to have anti-tumour effects through suppression of tumorigenesis, angiogenesis, and invasion. ATIC also resides in this module and is another potential fusion partner for ALK (Zhang *et al.*, 2010) (Chapter 1 Section 1.6.3), similar to EML4 (Blue/Green Yellow module).

Two genes with decreased expression over the course of cisplatin treatment in the green yellow module have previously used their down-regulation as a marker for disease and prognosis, or with poor survival (DPP4 [Javidroozi *et al.*, 2012] and POPDC3 [Luo *et al.*, 2012]). The adhesion molecule, JAM3, also in this module, has also had its expression correlated with progression in gastric adenocarcinoma (Hajjari *et al.*, 2013), while TGFBI has been associated with tumour suppression in mesothelioma (Li *et al.*, 2012) and MED29 has shown both oncogenic and tumour suppressive characteristics (Kuuselo *et al.*, 2011).

The purple module contains three genes previously seen in the automatic modules generated by the WGCNA algorithm (TNC, LPAR1, and CADM2) but also contains the cancer/testis antigen 45 A5 (CT45A5) that is thought to be a potential novel target for therapy (<http://ncbi.nlm.nih.gov/gene>) and TGFBR3, whose transient overexpression is thought to induce apoptosis in nasopharyngeal carcinoma cells (Zheng *et al.*, 2013).

Genes such as *EML4* and *EGFR* are already known to play roles in NSCLC. Others observed here that may have been seen in other cancers or are less well known in NSCLC, such as *CSAG2*, *ABCG2*, *PDXK*, *IGFBP7*, *HMOX1*, *ALX1*, *SPARC*, and *VSNL1*, may be key players in driving the cisplatin-resistant phenotype of the A549 cells.

Across all the modules of significance that have been focused on, there appear to be highly significant, differentially expressed genes that may be involved in oncogenic or tumour suppressor roles; that affect the adhesion status of the cell; initiate or propagate tumorigenesis, invasion, or metastasis; influence the apoptotic potential of the cells; are suggested to have involvement in EMT; are within key cancer-related pathways; or affect drug metabolism or efflux in to or out of the treated cells.

WGCNA also allows one to identify within modules potential key hub genes through measures of intra-modular connectivity, producing even more potential candidate driver genes. Some occur with high significance within modules, such as *EML4*, *DMKN*, and *CADM2* appearing in the IC25 dose 1 week recovery, independent of how the modules were created. In the IC50 dose 3 week recovery time course, even more hubs defined by high intra-modular connectivity appear in the top significantly differentially expressed genes, such as *MAL2*, *JAM3*, *UGT2B15*, *TP53INP1*, and *BMP7*. Other well described cancer genes such as *TP53* and *TP63* (Su *et al.*, 2012) are also seen as potential hubs in the two treated time courses. *SPARC* and *ITM2A* are two genes that are statistically suggestive of being hub genes with strong significant differential expression in both treated time courses, although the possible action of *ITM2A* is less well explored than *SPARC*.

Some of the other high intra-modular connectivity genes seen here have previously shown association with cancer phenotypes. *SLC6A14*, a member of the IC50 dose 3 week recovery time course blue module the latter correlating with IC50 value, could be a potential transporter of interest as it has already been related to cervical cancer (Gupta *et al.*, 2006), colorectal cancer and metastasis (Gupta *et al.*, 2005), and suggested as a potential drug target (Karunakaran *et al.*, 2011). *FOXQ1*, in the same blue module over expression is thought to influence NSCLC prognosis and associate with EMT (Feng *et al.*, 2012a) as well as promoting glioma cell proliferation and migration (Sun *et al.*, 2013).

Hub genes found only in the IC25 dose 1 week recovery include *NEDD9*, *GLI3*, and *ETS2*. *NEDD9* is associated with poor predictive prognosis of NSCLC (Kondo *et al.*, 2012) and may be capable of mediating lung cancer progression through LKB1 loss (Feng *et al.*, 2012b). This may link in with the *SGK1* gene discussed earlier (IC25 dose 1 week recovery blue module). *GLI3* is a hedgehog activator signal that enhances the tumour phenotype of

colorectal cancer through the up-regulation of adherence-related genes (Iwasaki *et al.*, 2012), which may help to explain the abundance of adherence genes seen in the differentially expressed genes. *ETS2* has been proposed as a potential candidate gene for squamous lung cancer (Bai and Hu, 2012) as well as possibly promoting etoposide resistance in cooperation with mutant p53 (Do *et al.*, 2012).

Investigations of the WGCNA generated modules for the 375 differentially expressed genes in common between the two time courses, but not present in the untreated time course, yielded further potential hubs, some of which can be found in the top genes from the individual time courses e.g. *JAM3* and *SCN9A*. Others have been associated with cancer too, such as *FGFR1*, *TAP1* and *SGCD*. *FGFR1* has had oncogenic amplifications in lung cancer documented (Oxnard *et al.*, 2013, Pros *et al.*, 2013) whilst overexpression of transporter, *TAP1*, has been associated with increased cancer cell-specific apoptosis and anti-tumour immune response in prostate cancer (Qiu *et al.*, 2012). An intronic SNP within the *SGCD* gene has been associated with increased expression of *SGCD* and subsequent paclitaxel sensitivity in the NCI60 cancer cell line (Eng *et al.*, 2011).

In amongst all these potential genes could be key mutated driver genes that encourage the tumour progressive or cisplatin-resistant phenotype seen in the cultured cells over the time course (Chapter 3 Section 3.4). The recurrence of well-known cancer genes appearing as potential hubs from the WGCNA analyses suggest that the cancer or drug resistant phenotype investigated here, though obviously heterogeneous, may be as a result of previously characterised signalling pathways, for which the GO analyses provides some evidence. It is possible that the mechanisms of resistance to platinum compounds and other chemotherapeutics are general within cancer cells. To reduce the numbers of genes and validate potential biomarkers, comparison of these results with differential gene expression

data from patients, both of normal and tumour tissue, especially those of treated versus untreated with platinum chemotherapy (Chapter 4 Section 4.5.2), will be vital.

## 6 Gene Expression Analysis of Patient Samples

### 6.1 The Need for Validation in Patient Samples

The *in vitro* cellular model of cisplatin dosing that has been optimised and established (Chapter 3 Sections 3.2-3.4) has generated, through transcriptome studies, a number of gene lists of potential interest in relation to cisplatin resistance (Chapters 4 and 5). These genes now require investigation in an *in vivo* setting in order to validate their potential as indicators of changing genomic response to platinum compounds. If any show substantial and significant differences between platinum-treated and untreated patients, it is possible they may be useful markers of response to platinum or be helpful in the elucidation of the pathways of evasion from cisplatin-induced death.

Two cohorts of patients have been used for this validation but there are some important factors that have to be considered when working with whole tissue samples. Perhaps the most critical is the recognition that samples are heterogeneous in nature. It is feasible that the biopsy taken is not representative of the tumour, or normal lung tissue, as a whole. The fact that all the patients are individuals with their own genetic backgrounds may also lead to strong inherent variability between the samples that are subsequently grouped together for analyses. Tumours may appear to have mixed morphologies too that may alter their genomics and lead to different profiles. This may also be the case with tumours at different stages or with different nodal involvements and therefore careful phenotyping is essential.

The use of different surgical approaches may lead to the potential for some variability in the samples, such as biopsy from lung resection versus biopsy from rigid bronchoscopy, which may also lead to differences in the genomics of the tissue. In addition surgeons may be

limited spatially as to what tissue is available for biopsy. Areas of necrosis and fibrosis must also be avoided in an attempt to acquire an adequate viable sample of the tumour.

The selection of matched “normal” tissue from an individual is difficult to choose; it is possible, if not probable that the area surrounding the tumour is not equivalent to true normal lung due to its proximity to the tumour itself and so may phenotypically be closer to a neoplastic phenotype. “Normal” samples may even be of a totally different structure such as fat tissue, and not be representative of the non-neoplastic lung. If a large sample is taken that is then dissected into smaller samples for research, this adds another level of potential sampling error. Small sample sizes are equally a problem as their representativeness is questionable and downstream processing, such as physical handling, and RNA and DNA extraction is technically more challenging.

The freezing of the sample after biopsy may affect tissue integrity. As little delay as possible between excisions and freezing should be attempted to ensure that the state of the tumour is as similar to *in situ* as possible. The freezing process should be rapid to prevent any large crystal formation within the sample that could affect the cellular structure.

In spite of these various limitations and potential pitfalls, human samples give us the opportunity to look at what is actually happening in human lung cancers. Tissue samples from two patient cohorts, RESTORE-AIR and EQUALITY, both ascertained and recruited through St George’s Hospital London, were available to explore further the WCGNA and expression clustering findings from the *in vitro* cellular model of cisplatin resistance.

## **6.2 RESTORE-AIR – Randomised Evaluation of STents to Open REstricted AIRways in patients with centrally placed non-small cell lung cancer**

The RESTORE-AIR study builds upon a previous pilot study that had shown good improvement in breathlessness in lung cancer patients with the insertion of a bronchial stent. Patients with a NSCLC diagnosis and a central tumour suitable for rigid bronchoscopy and bronchial stenting were evaluated for study inclusion.

Patients were consented for voluntary tissue biopsy, by rigid bronchoscopy, at time of stent insertion. If consent was given, three sets of biopsies were taken under general anaesthetic whenever possible. The first set of tumour and normal bronchial tissue biopsies were immediately placed in RNALater [Qiagen] (RNAse inactivation buffer). The second pair were frozen in liquid nitrogen and subsequently stored at -80°C, whilst the final pair were used to make formalin-fixed, paraffin embedded (FFPE) blocks for immuno-histochemical staining by the pathology laboratory.

A total of eight paired samples stored in RNALater were available for the investigations done as part of this thesis. Demographics for the seven patients are given in Table 6.1. Seven out of eight patients had been treated with platinum and the majority (six out of seven) of these had relapsed. It is worth noting that sample 203 had been exposed to many more lines of drug therapy than any other (eight compared to zero to two). This may have had a marked effect on the comparative gene expression profiles and genome in general after many potentially cytotoxic drug exposures.

| ID  | Age | Smoker? | Ethnicity     | Relapsed from time of diagnosis? | No. of Previous Lines of Drug Therapy | Previous Platinum treatment? | Platinum -free Interval (months) |
|-----|-----|---------|---------------|----------------------------------|---------------------------------------|------------------------------|----------------------------------|
| 101 | 74  | Current | White British | Y                                | 2                                     | Y                            | 12                               |
| 203 | 81  | Ex      | White British | Y                                | 8                                     | Y                            | 19                               |
| 204 | 67  | Ex      | White British | Y                                | 1                                     | Y                            | 7                                |
| 206 | 59  | Ex      | White British | Y                                | 1                                     | Y                            | 21                               |
| 207 | 63  | Current | White British | Y                                | 2                                     | Y                            | 17                               |
| 208 | 61  | Current | White British | N                                | 0                                     | N                            | 0                                |
| 214 | 61  | Ex      | White British | Y                                | 1                                     | Y                            | 4                                |
| 215 | 70  | Never   | White British | N                                | 1                                     | Y                            | 5                                |

**Table 6.1** Demographics for the seven RESTORE-AIR study participants. Details include exposure to and relapse after chemotherapy treatment with platinum compounds. Seven patients have been treated with platinum and six of these have relapsed. Sample 203 has been exposed to many more lines of drug therapy than any other (eight compared to zero to two). All samples were collected by rigid bronchoscopy.

### **6.3 EQUALITY – Expression QUantitative trait Loci mapping In Tumours collected at biopsy**

EQUALITY is an on-going prospective study to identify expression quantitative trait loci in patients with thoracic cancers undergoing biopsy. Patients, referred to St George’s Hospital for consideration of clinically indicated biopsy by rigid bronchoscopy or thoracic surgery, with a likely or proven diagnosis of NSCLC, SCLC, mesothelioma, or pulmonary carcinoid have been recruited (nine in total at the time of this current study). Only those of subtypes of NSCLC i.e. adenocarcinomas and squamous cell carcinomas were included for this study. During rigid bronchoscopy or thoracic surgery, three sets of research biopsies were taken

when feasible, with processing and storage identical to that used for the RESTORE-AIR. In addition, at the time of registration, samples of the participants' blood, plasma and urine were taken for future genetic, metabolomic and proteomic work.

Only four of the nine EQUALITY samples collected, all obtained by rigid bronchoscopy, were suitable (NSCLC adenocarcinoma or squamous histological subtypes) to be combined with those from the RESTORE-AIR study for the validation work. All available demographics are given in Table 6.2. Details such as relapse after therapy and platinum-free interval were unavailable.

| ID  | Age | Subtype        | Smoker? | Ethnicity     | Previous Platinum treatment? | No. of Previous Lines of Drug Therapy |
|-----|-----|----------------|---------|---------------|------------------------------|---------------------------------------|
| EQ3 | 70  | Squamous       | Ex      | White British | Y                            | 1                                     |
| EQ4 | 82  | Squamous       | Ex      | White British | N                            | 0                                     |
| EQ6 | 52  | Squamous       | Ex      | White British | Y                            | Unknown                               |
| EQ7 | 60  | Adenocarcinoma | Ex      | White British | Y                            | 1                                     |

**Table 6.2** Demographics for EQUALITY patients that can be combined with those from the RESTORE-AIR study for these investigations.

## 6.4 RNA Extraction from Tissue Samples

RNA was extracted from patient tissue samples using the QIAgen RNeasy Fibrous Tissue Mini kit with a few minor modifications to the manufacturer's instructions in order to optimise yield (Chapter 2 Section 2.2.3.1). To improve disruption and homogenisation of the tissue rather than conducting a single homogenisation of 20-40s with the TissueRuptor, several 30s homogenisations were done up to a maximum of five with samples being placed on ice in-between bursts. In addition the length of the Proteinase K digestion was increased from 10 minutes to two hours. All resulting RNAs had RINs  $\geq 8.5$ .

Despite these optimisations it was observed that normal non-tumour tissue in general tended to result in lower yields of RNA, insufficient for microarray work, compared to tumour tissues, although a few of the latter gave miniscule yields. This difference (normal versus tumour tissue yield) could be due to intrinsic differences in the activity of tissues, such as some tumours may be more highly genomically active, giving larger yields of RNA than their normal tissue counterparts, or there may be inherent variability between some tissues and others leading to the wide variation seen in RNA concentrations.

A final contributing factor might be the fact that RESTORE-AIR is an historic collection and the recommended protocol for storing tissue in the RNA stabilisation reagent RNALater has since been modified from when the original collection took place and the present time. Originally tissue was stored in the RNALater at -20°C over the course of 24 hours before being transferred to -80°C for long term storage. Recently the manufacturers (Qiagen – suppliers of RNALater and the RNeasy fibrous kit) recommend removing the stabilisation reagent before transfer of samples to -20°C/-80°C for archival storage. This suboptimal procedure combined with a relatively long freeze (at least four years) could have potentially led to some of the reduced yields observed upon extraction. The same procedure was applied for the EQUALITY sampling.

In total, nine tumour and six normal samples were successfully extracted (Table 6.3) and subject to processing for Affymetrix Human Gene 1.1 ST arrays (Chapter 2 Section 2.2.3.2). Four normal (RA203N, RA204N, EQ3N, and EQ4N) and three tumour samples (RA101T, RA204T, and EQ7T) gave too low RNA yield after extraction for further global gene expression analyses.

| Study       | Sample | Subtype        | Tumour/<br>Normal | Concentration<br>(µg/µl) |
|-------------|--------|----------------|-------------------|--------------------------|
| RESTORE-AIR | 203    | Squamous       | T                 | 97.2                     |
| RESTORE-AIR | 206    | N/A            | N                 | 47.6                     |
| RESTORE-AIR | 206    | Squamous       | T                 | 2787.2                   |
| RESTORE-AIR | 207    | N/A            | N                 | 186.7                    |
| RESTORE-AIR | 207    | Squamous       | T                 | 58.9                     |
| RESTORE-AIR | 208    | N/A            | N                 | 338.8                    |
| RESTORE-AIR | 208    | Squamous       | T                 | 72.6                     |
| RESTORE-AIR | 214    | N/A            | N                 | 227.9                    |
| RESTORE-AIR | 214    | Squamous       | T                 | 1273.6                   |
| RESTORE-AIR | 215    | N/A            | N                 | 75.1                     |
| RESTORE-AIR | 215    | Adenocarcinoma | T                 | 60.8                     |
| EQUALITY    | 3      | Squamous       | T                 | 48.2                     |
| EQUALITY    | 4      | Squamous       | T                 | 326.8                    |
| EQUALITY    | 6      | N/A            | N                 | 310.5                    |
| EQUALITY    | 6      | Squamous       | T                 | 1150.9                   |

**Table 6.3** RNA extraction from tissue for RESTORE-AIR and EQUALITY samples showing the resulting concentrations. Four normal and three tumour samples gave too low RNA yield after extraction for further global gene expression analyses. The remaining fifteen were subject to RNA processing and microarray hybridisation for gene expression analyses.

## 6.5 A Dataset for Comparisons of Platinum Treatment versus Non-Treated Samples

Data from a set of non-platinum-treated tumour samples was available (Molecular Genetics Group: unpublished data) was used as a control for the platinum-treated patient tumour tissues. These 35 control samples were taken at the time of surgical resection. The dataset consists of global gene expression profiles of paired (normal and tumour) tissue samples, either adenocarcinoma (20 samples) or squamous cell carcinoma (15 samples). The untreated samples selected for comparative analysis from this extensive dataset were matched as

closely as possible for age, gender, smoking status (Current, Ex or Never) and subtype at the time of biopsy, to those from the EQUALITY and RESTORE-AIR studies. Careful batch removal and outlier detection was required to ensure any false positives and negatives were eliminated as far as possible. All patients were of white British or Irish ethnicity. For ease of analyses, all adenocarcinoma samples (including RA215T) were removed to retain focus on the squamous subtype.

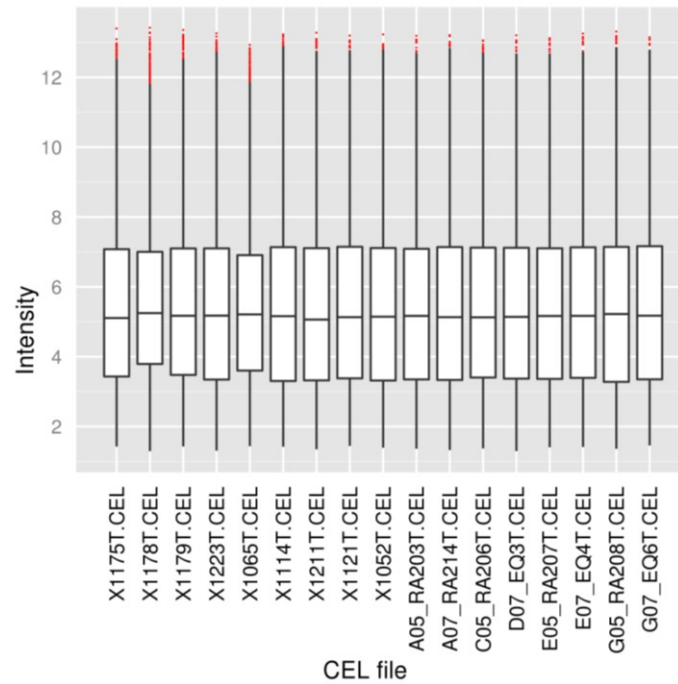
#### 6.5.1 Outlier Detection and Removal, and Microarray Quality Control

The selected six cases (platinum-treated) and eleven out of a possible fourteen controls (non-platinum-treated) (Table 6.4) were subject to the same quality control measures as previous CEL files (details given in Chapter 4 Section 4.2) prior to differential expression analysis. Briefly the seventeen CEL files for the proposed matches between cases and controls were subject to RMA normalisation using APT, followed by stringent outlier removal using inter-array correlation (IAC) – the Pearson’s correlation coefficient of the expression levels for a given pair of microarrays using all probe sets for which data are available.

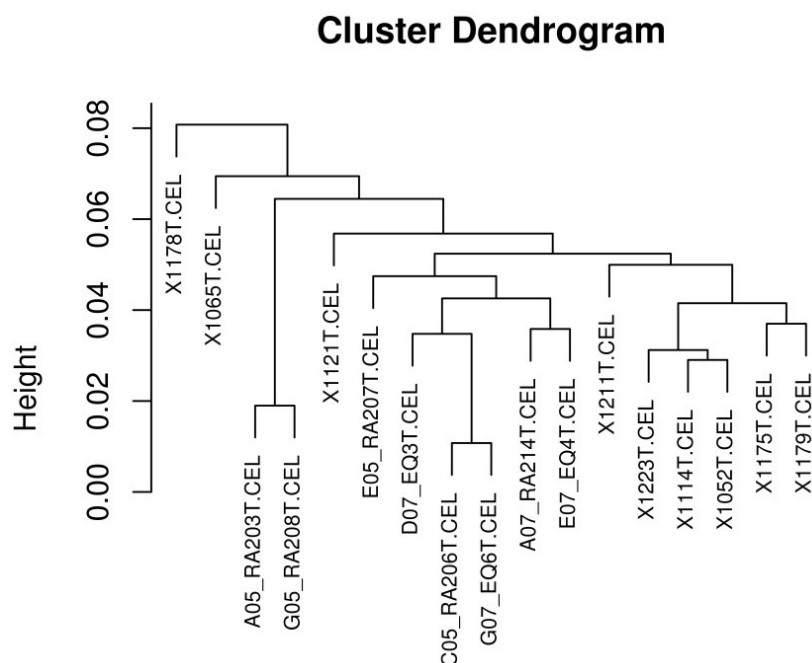
The removal of outliers is extremely important, especially in investigations where a historical dataset is being used for comparison. IAC outlier exclusion is based solely on microarray samples with divergent gene expression levels, with no operator bias or input and so produces a dataset with overall consistent and comparable of gene expression amongst all the samples used. The RMA box plot of the seventeen CEL files used for these analyses (Figure 6.1) infers that some of these samples may be strong outliers, a fact which was confirmed after rounds of IAC (Figure 6.2 and 6.3). Any sample with a mean IAC more than two standard deviation of away from the total mean IAC was therefore excluded as an outlier and successive rounds of this logical sample removal strategy were implemented until none of the CEL files fitted this criterion (Figure 6.4).

| Study                | ID   | Age | Gender | Histology | Pt Treatment | Smoker  | Ethnicity     |
|----------------------|------|-----|--------|-----------|--------------|---------|---------------|
| EQUALITY             | 3    | 70  | F      | Squamous  | Y            | Ex      | White British |
| EQUALITY             | 4    | 82  | M      | Squamous  | N            | Ex      | White British |
| EQUALITY             | 6    | 52  | F      | Squamous  | Y            | Ex      | White British |
| RESTORE-AIR          | 203  | 81  | M      | Squamous  | Y            | Ex      | White British |
| RESTORE-AIR          | 206  | 59  | M      | Squamous  | Y            | Ex      | White British |
| RESTORE-AIR          | 207  | 63  | M      | Squamous  | Y            | Current | White British |
| RESTORE-AIR          | 208  | 61  | M      | Squamous  | N            | Current | White British |
| RESTORE-AIR          | 214  | 61  | M      | Squamous  | Y            | Ex      | White British |
| Lung Resection Study | 1052 | 66  | F      | Squamous  | N            | Ex      | White British |
| Lung Resection Study | 1065 | 63  | M      | Squamous  | N            | Ex      | White British |
| Lung Resection Study | 1114 | 71  | F      | Squamous  | N            | Ex      | White British |
| Lung Resection Study | 1119 | 74  | M      | Squamous  | N            | Ex      | White British |
| Lung Resection Study | 1121 | 78  | M      | Squamous  | N            | Ex      | White Irish   |
| Lung Resection Study | 1175 | 82  | M      | Squamous  | N            | Ex      | White British |
| Lung Resection Study | 1178 | 82  | M      | Squamous  | N            | Ex      | White British |
| Lung Resection Study | 1179 | 68  | M      | Squamous  | N            | Current | White British |
| Lung Resection Study | 1186 | 81  | F      | Squamous  | N            | Never   | White British |
| Lung Resection Study | 1199 | 76  | M      | Squamous  | N            | Ex      | White British |
| Lung Resection Study | 1206 | 63  | F      | Squamous  | N            | Current | White British |
| Lung Resection Study | 1209 | 68  | F      | Squamous  | N            | Current | White British |
| Lung Resection Study | 1211 | 71  | F      | Squamous  | N            | Ex      | White British |
| Lung Resection Study | 1223 | 68  | M      | Squamous  | N            | Current | White British |

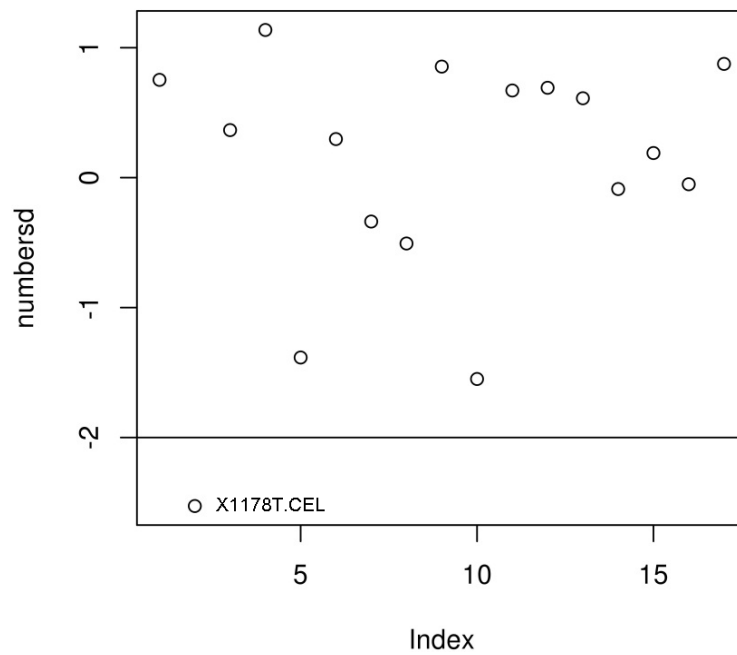
**Table 6.4** Demographics for Patient Samples for differential expression analyses. Samples 1052, 1065, 1114, 1121, 1175, 1178, 1179, 1211, 1223 and RA208 were used as controls to match the six platinum-treated patients (RA203, RA206, RA207, EQ3, EQ4, and EQ6) for age, gender, and smoking status as closely as possible. All patients were of white British or Irish ethnicity and any adenocarcinoma cases were excluded to prevent possible batch effects.



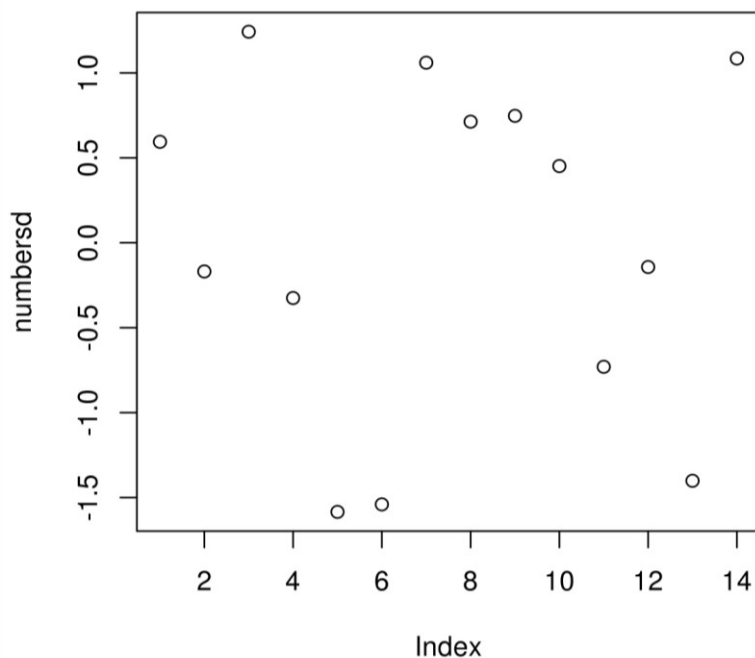
**Figure 6.1** Box plot of normalised intensities (log expression signal) of the seventeen patient samples. X1065T.CEL and X1178T.CEL appear smaller than the other boxes and with a greater number of potential outlier points (red dots at the end of the whisker of the plot).



**Figure 6.2** Hierarchical cluster dendrogram of all seventeen .CEL files using 1-IAC as the distance metric. It can be seen that X1065T.CEL and X1178T.CEL appear again as outliers on this plot, as observed in **Figure 6.1**.

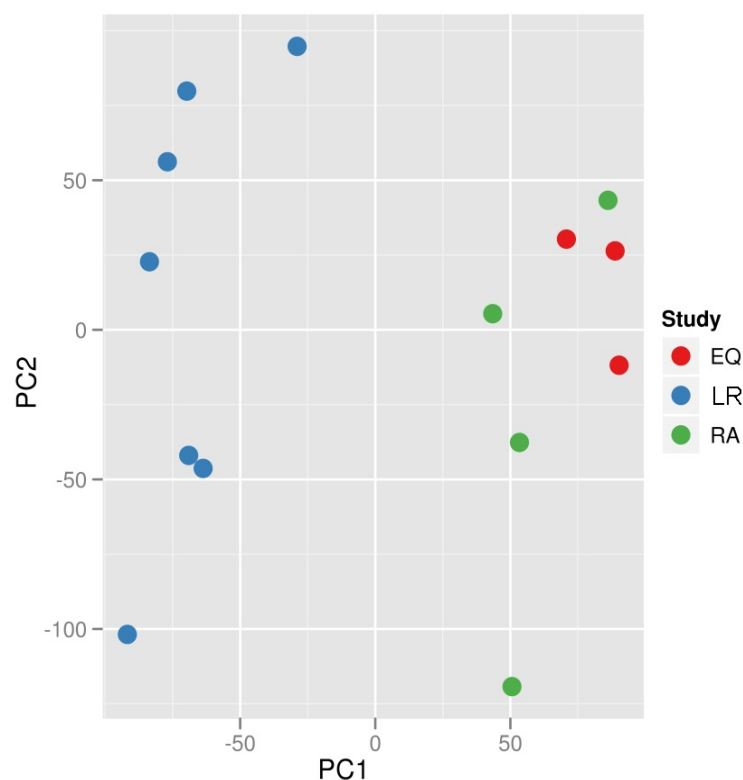


**Figure 6.3** Plot showing the distribution of the mean IAC for each array, where numbersd equals (the array mean IAC – mean of all mean IAC) divided by the standard deviation of the mean IAC. An outlier was defined as being at least two standard deviations outside the mean of all mean IAC. Here it can be seen that one sample (X1178T.CEL) fulfils this criteria and was subsequently removed. A further round of this outlier removal led to X1065T.CEL and RA203T.CEL also being excluded.



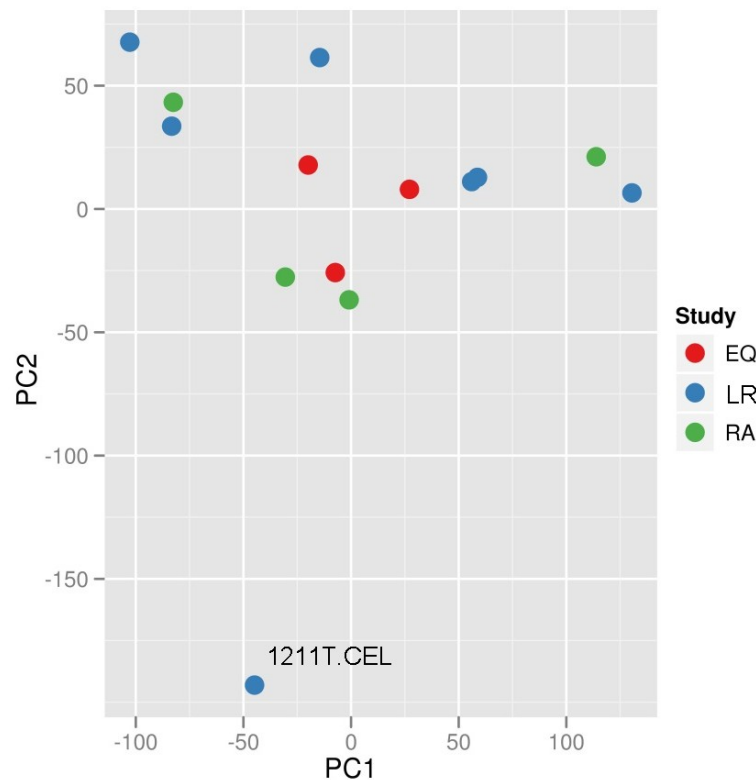
**Figure 6.4** Plot showing the distribution of mean IAC for each array after outlier removal with no CEL files falling outside the standard deviation cut off. All these remaining fourteen samples were taken into PCA.

The method of IAC outlier exclusion led to the removal of 1178T.CEL, 1065T.CEL and RA203T.CEL and the remaining fourteen samples were subject to PCA to check for potential batch effects in this reduced dataset. As can be seen in Figure 6.5, there is a definite split between the samples from the lung resection study (blue points, legend = LR) and the EQUALITY and RESTORE-AIR studies (red and green points, legend = EQ and RA respectively). This is most likely a reflection of the different methods of tissue sampling, surgical lung resection versus biopsy by rigid bronchoscopy. The RNA extraction methods and array platforms implemented for all samples were the same. Fortunately this split can be removed thanks to the presence of untreated controls in the rigid bronchoscopy studies, thus differences due to treatment will be retained.



**Figure 6.5** PCA plot showing the distribution of the three studies. It can be seen that the red and green (EQUALITY (EQ) and RESTORE-AIR (RA) samples) cluster together on the right hand side of the plot while the blue (Lung Resection (LR) Study) cluster on the left of the plot. Thanks to the presence of untreated samples in both clusters this batch effect, potentially caused by the two different methods of biopsy, can be corrected for.

After correction for this strong batch effect in the RMA data another PCA was performed. Batch correction led to a more even scattering of the samples across the PCA plot (Figure 6.6) although one single outlier (1211T.CEL) still remained and was therefore excluded. This strong batch effect was removed in the study design for linear modelling before analysis of differential expression between treated and non-treated tumours took place.



**Figure 6.6** PCA plot showing how the batch effect of study has been removed to give a more even dataset. Only one outlier 1211T.CEL remained and was subsequently removed and excluded from further analyses. The same legend applies as for **Figure 6.5**, i.e. red are for EQUALITY (EQ) samples, green are for RESTORE-AIR (RA) samples, and blue are for the Lung Resection Study (LR).

Due to the mixture of three different studies, RESTORE-AIR, EQUALITY and the Lung Resection Study, all of which originally had different purposes, the numbers and matches of treated and untreated patients has not been easy. The presence of an untreated sample within both rigid bronchoscopy data sets (RESTORE-AIR and EQUALITY) is fortunate as it

allowed the strong batch effect of “Study” (Figure 6.5) to be controlled and the data to be compared.

The only platinum-treated sample to be removed by outlier detection was RA203T; the only patient to have received more than two lines of previous drug therapy. It is possible that the many rounds of successive therapies caused gross genetic changes to the patient’s tumour, and possibly normal lung tissue. The tumour may also have evolved naturally over time, as the patient may have lived longer than the others in order to receive these therapies. These may have been so profound that their expression profile was extensively different from the other samples and could have led to its outlier status.

## **6.6 Differential Expression Analysis**

The RMA data from the thirteen samples that had passed quality control and outlier exclusion (Section 6.5.1) were used to explore the differences in expression of the genes from the lists and modules of interest generated from the cisplatin resistance cell model (Chapters 4 and 5). These gene lists of interest were the increasing and decreasing expression clusters from the IC50 dose 3 week recovery time course and the IC25 dose 1 week recovery time course (Chapter 4 Sections 4.4.1.1 and 4.4.2.1) and modules of interest comprised of those with strong positive and negative correlations with IC value of cisplatin (Chapter 5 Sections 5.2.3, 5.2.4 and 5.2.6) (Table 6.5).

Each individual gene list, ranging between 31 and 323 genes, was subject to a separate differential expression analysis, conducted using the RMA data from the patient samples comparing those platinum-treated cases with untreated control patients. Study, gender, and smoking status were all included as batch effects in the linear model performed using the R

package Limma. The additional variable of the age of patient could not be included, as the linear model could not be generated with few individuals being of exactly the same age (range=52-82, mean=70, 14 unique ages, Table 6.4). Each treated sample was contrasted with every untreated sample to give a T-statistic for each gene and the resultant  $p$ -value was B-H adjusted for multiple comparisons, as with all previously reported statistics. Use of these gene lists reduces the number of comparisons performed in one analysis and hence should yield statistically significant and stringent  $p$ -values. With this in mind, the threshold applied for significance was B-H adjusted  $p$ -value  $<0.01$ . Most clusters or modules harboured genes that were significantly differentially expressed between the treated and untreated samples with twenty one genes with B-H adjusted  $p$ -values  $<0.01$  with a number exhibiting substantial fold changes in expression (Table 6.6).

| Time Course               | Module or Cluster                                              | Number of Genes |
|---------------------------|----------------------------------------------------------------|-----------------|
| IC50 Dose 3 Week Recovery | Increasing expression                                          | 227             |
| IC50 Dose 3 Week Recovery | Decreasing expression                                          | 209             |
| IC25 Dose 1 Week Recovery | Increasing expression                                          | 285             |
| IC25 Dose 1 Week Recovery | Decreasing expression                                          | 323             |
| IC50 Dose 3 Week Recovery | Blue Module – Positive<br>Correlation with IC50 value          | 278             |
| IC50 Dose 3 Week Recovery | Pink Module – Positive<br>Correlation with IC50 value          | 50              |
| IC50 Dose 3 Week Recovery | Green Module – Negative<br>Correlation with IC50 value         | 150             |
| IC50 Dose 3 Week Recovery | Yellow Module – Negative<br>Correlation with IC50 value        | 152             |
| IC25 Dose 1 Week Recovery | Midnight Blue Module – Negative<br>Correlation with IC25 Value | 31              |
| IC25 Dose 1 Week Recovery | Green Module – Negative<br>Correlation with IC25 Value         | 138             |
| IC25 Dose 1 Week Recovery | Green Yellow Module – Positive<br>Correlation with IC25 Value  | 49              |
| IC25 Dose 1 Week Recovery | Purple Module – Negative<br>Correlation with IC25 Value        | 65              |
| IC25 Dose 1 Week Recovery | Light Cyan Module – Negative<br>Correlation with IC25 Value    | 25              |
| IC25 Dose 1 Week Recovery | Blue Module – Negative<br>Correlation with IC25 Value          | 171             |
| IC25 Dose 1 Week Recovery | Salmon Module – Positive<br>Correlation with IC25 Value        | 40              |
| 375 Overlapping Genes     | Green Module – Positive<br>Correlation with IC50 Value         | 41              |
| 375 Overlapping Genes     | Yellow Module – Positive<br>Correlation with IC50 Value        | 46              |
| 375 Overlapping Genes     | Blue Module – Positive<br>Correlation with IC50 Value          | 64              |

**Table 6.5** The time course expression clusters and WGCNA modules, including the number of genes per list, used to perform differential expression analyses between the treated and untreated patient tissues. Positive and negative correlations with respective IC50 or IC25 value have been indicated.

**Table 6.6** Top 21 Genes, seen in patients, from differential expression analyses using the genes of interest taken from the *in vitro* cellular model. From <http://www.ncbi.nlm.nih.gov/gene> and <http://www.genecards.org> unless otherwise stated.

| Gene                | B-H adjusted <i>p</i> -value                                               | Fold change (3 s.f.) | Expression Cluster | WGCNA Module of interest                                              | Function                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
|---------------------|----------------------------------------------------------------------------|----------------------|--------------------|-----------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b><i>GAL</i></b>   | 0.0000270 (IC25 Salmon)<br>0.0000500 (Overlap Blue)<br>0.000233 (3wk Blue) | 0.261                |                    | IC25 Salmon Module<br><br>Overlap Blue Module<br><br>3wk Blue Module  | Galanin-GMAP prepropeptide – a small neuropeptide that functions as cellular messenger in central and peripheral nervous system. It contracts the smooth muscle of gastrointestinal and genitourinary tract, regulates growth hormone release, modulates insulin release, and the control of adrenal secretion.<br><br>Disruption of galanin expression or galanin receptor signalling has been seen in many multifactorial conditions suggesting role in development and/or pathology of certain diseases, including cancer. It and its receptor GAI1 have been identified as novel determinants of chemotherapy resistance and potential biomarkers in colorectal cancer (Stevenson <i>et al.</i> , 2012). |
| <b><i>INA</i></b>   | 0.00201 (Overlap Blue)<br>0.00563 (3wk Down)<br>0.00922 (3wk Blue)         | 10.4                 | 3wk Down Cluster   | Overlap Blue Module<br><br>3wk Blue Module                            | Internexin – neuronal intermediate filament protein alpha. It covers the axoskeleton and functionally maintains neuronal calibre. It may play a role in intracellular transport to axons and dendrites and be involved in neuronal morphogenesis.<br><br>Immunohistochemical detection of IDH1 mutation, p53 and internexin as prognostic factors of glial tumours (Takano <i>et al.</i> , 2012).                                                                                                                                                                                                                                                                                                            |
| <b><i>SCN9A</i></b> | 0.00594 (Overlap Yellow)<br>0.00843 (3wk Pink)<br>0.00941 (IC25 Green)     | 7.47                 |                    | Overlap Yellow Module<br><br>3wk Pink Module<br><br>IC25 Green Module | Sodium Channel voltage-gated type IX alpha subunit.<br><br>Significant role in nociception signalling mutations associated with primary erythromalgia, channelopathy-associated insensitivity to pain.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |

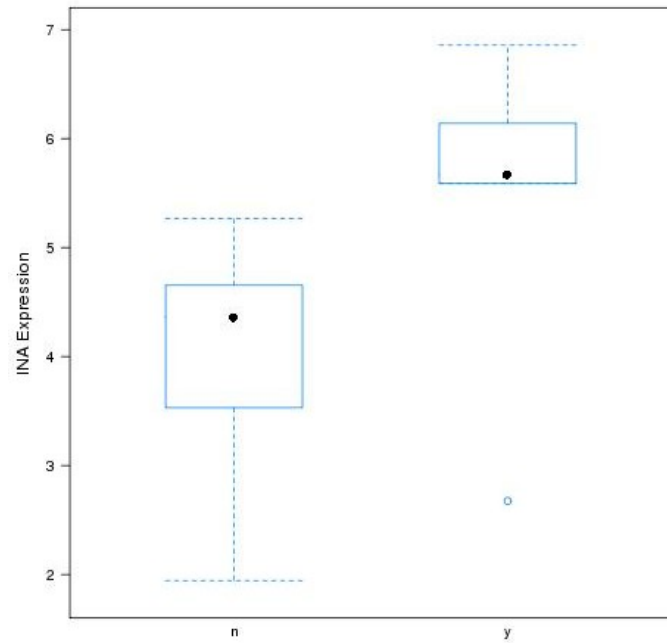
|                       |                                                              |       |                   |                          |                                                                                                                                                                                                                                                                                                                        |
|-----------------------|--------------------------------------------------------------|-------|-------------------|--------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b><i>ABHD4</i></b>   | 0.000356<br>(3wk Blue)<br><br>0.000393<br>(3wk Up)           | 2.69  | 3wk Up Cluster    | 3wk Blue Module          | $\alpha/\beta$ hydrolase domain containing 4.<br><br>A genome-wide short-hairpin RNA screen identified ABHD4 as a potential novel regulator of resistance to apoptosis induced by anchorage-dependent cells detaching from the extracellular matrix (anoikis) (Simpson <i>et al.</i> , 2012).                          |
| <b><i>C3orf57</i></b> | 0.00420<br>(IC25 Down)<br><br>0.000499<br>(IC25 Salmon)      | 3.26  | IC25 Down Cluster | IC25 Salmon Module       | Alternative gene name SPTSSB – Serine palmitoyltransferase small subunit B. It catalyses the first committed and rate-limiting step in sphingolipid synthesis and may play a role in signal transduction.                                                                                                              |
| <b><i>CYP2S1</i></b>  | 0.00278<br>(3wk Down)<br><br>0.00292<br>(3wk Yellow)         | 2.99  | 3wk Down Cluster  | 3wk Yellow Module        | Cytochrome P450 family 2 subfamily S polypeptide 1 – a monooxygenase, which catalyses many reactions involved in drug metabolism and synthesis of cholesterol, steroids and other lipids. Its homologous protein in rodents has been shown to metabolise carcinogens but, in humans, the specific function is unknown. |
| <b><i>HMCN1</i></b>   | 0.00539<br>(3wk Blue)<br><br>0.00643<br>(3wk Down)           | 4.13  | 3wk Down Cluster  | 3wk Blue Module          | Hemicentin1 – a large extracellular member of immunoglobulin super-family.                                                                                                                                                                                                                                             |
| <b><i>RAB15</i></b>   | 0.00539<br>(3wk Blue)<br><br>0.00805<br>(3wk Up)             | 2.09  | 3wk Up Cluster    | 3wk Blue Module          | RAB15 – a member of the KRAS signalling pathway.<br><br>Expression has been shown to correlate with retinoic acid-induced differentiation of neuroblastoma cells (Nishimura <i>et al.</i> , 2011).                                                                                                                     |
| <b><i>UTY</i></b>     | 0.00315<br>(IC25 Green Yellow)<br><br>0.00794<br>(IC25 Down) | 0.243 | IC25 Down Cluster | IC25 Green Yellow Module | Ubiquitously transcribed tetratricopeptide repeat gene Y-linked – is a potential histone de-methylase. It contains tetratricopeptide repeats that are thought to be involved in protein-protein interactions.                                                                                                          |

|                        |         |        |                   |                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|------------------------|---------|--------|-------------------|--------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b><i>AFAP1</i></b>    | 0.00198 | 0.0804 |                   | 3wk Blue Module          | <p>Actin filament associated protein 1 – an SRC binding partner and potential modulator of actin filament integrity in response to cell signals. It may function as adaptor protein by linking SRC family members and/or other signalling proteins to actin filaments.</p> <p>It has been potentially shown to play role in development and progression of prostate adenocarcinoma by regulating cell-matrix adhesions and migration in cancer cells.</p>                                                                                                                                                                                                                                                                |
| <b><i>CYorf15A</i></b> | 0.00316 | 0.277  |                   | IC25 Green Yellow Module | TXLNG2P – Taxilin Gamma 2 Pseudogene.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| <b><i>KDM5D</i></b>    | 0.00315 | 0.209  |                   | IC25 Green Yellow Module | Lysine-specific demethylase 5D Y-linked.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| <b><i>MAP2K3</i></b>   | 0.00420 | 2.80   | IC25 Down Cluster |                          | <p>Mitogen activated protein kinase kinase 3 – a dual specificity protein kinase that can be activated by insulin and by cytokines/environmental stress. It catalyses concomitant phosphorylation of threonine and tyrosine residue in MAP kinase p38.</p> <p>Proteome profiling of immortalisation to senescence transition of human breast epithelial cells identified MAP2K3 as a senescence-promoting protein, which is down-regulated in human breast cancer (Jia <i>et al.</i>, 2010).</p> <p>RAS oncogene expression was found to result in the accumulation of this kinase's active form, subsequently leading to constitutive activation of MAPK14 and conferred oncogenic transformation of primary cells.</p> |
| <b><i>MGC70870</i></b> | 0.00941 | 0.348  |                   | IC25 Green Module        | C-terminal binding protein 2 pseudogene.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| <b><i>QRFP</i></b>     | 0.00760 | 5.53   |                   | 3wk Blue Module          | <p>Pyroglutamylated RFamide peptide receptor – receptor for the orexigenic neuropeptide QRFP.</p> <p>Its activity is mediated by G proteins that modulate adenylate cyclase activity and intracellular calcium levels.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |

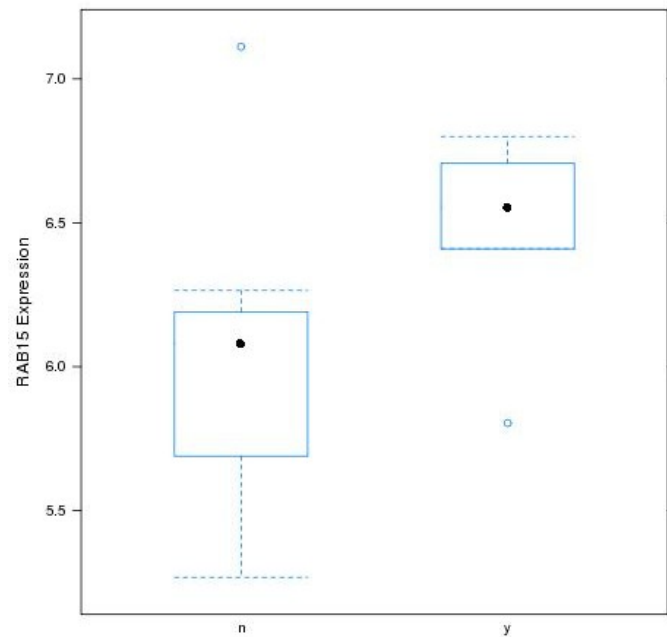
|                       |          |       |                   |                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|-----------------------|----------|-------|-------------------|--------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b><i>RGS2</i></b>    | 0.000286 | 2.71  |                   | IC25 Salmon Module       | <p>Regulator of G-protein signalling - acts as GTPase activating proteins for G alpha subunits of heterotrimeric G proteins and drive G proteins into inactive GDP-bound forms. It has roles in leukemogenesis and the negative feedback control pathway for adenylyl cyclase signalling. It also binds EIF2B5 and blocks its activity inhibiting translation of mRNA into protein.</p> <p>In human airway epithelial cells, RGS 2 has been shown to inhibit acid-induced mucin5AC hypersecretion, and its relatives RGS10 and 17 regulate chemoresistance in ovarian cancer cells (Liu <i>et al.</i>, 2013, Sethakorn and Dulin, 2013).</p> |
| <b><i>RPS4Y1</i></b>  | 0.000140 | 0.166 |                   | IC25 Green Module        | <p>Ribosomal protein S4 Y-linked 1.</p> <p>Patient gender is associated with distinct patterns of chromosomal abnormalities and sex chromosome linked gene-expression profiles in meningiomas.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| <b><i>SERINC5</i></b> | 0.00420  | 1.95  | IC25 Down Cluster |                          | <p>Serine Incorporator 5.</p> <p>Enhances the incorporation of serine into phosphatidylserine and sphingolipids.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| <b><i>SLC46A3</i></b> | 0.00642  | 0.500 |                   | 3wk Blue Module          | Solute carrier family 46 member 3.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| <b><i>ST3GAL1</i></b> | 0.00407  | 2.19  |                   | IC25 Green Module        | <p>ST3 beta-galactoside alpha-2,3-sialyltransferase – a type II membrane protein that catalyses the transfer of sialic acid from CMP-sialic acid to galactose-containing substrates.</p> <p>Transcript levels of ST3GAL1 and other glycosyltransferase genes were significantly elevated in colon cancer cells undergoing EGF/bFGF-induced epithelial-mesenchymal transition, possibly through the action of known cancer gene, cMYC (Sakuma <i>et al.</i>, 2012).</p>                                                                                                                                                                       |
| <b><i>USP9Y</i></b>   | 0.00657  | 0.283 |                   | IC25 Green Yellow Module | Ubiquitin specific peptidase 9, Y-linked – a member of peptidase C19 family that may play an important regulatory role at protein turnover level. It is an essential component of TGFbeta/BMP signalling cascade and deubiquitinates monoubiquitinated SMAD4, opposing the activity of E3 ubiquitin-protein ligase TRIM33.                                                                                                                                                                                                                                                                                                                   |

The genes *INA*, *SCN9A*, *ABHD4*, *C3orf57*, *CYP2S1*, *RAB15*, *HMCN1*, *MAP2K3*, *QRFPR*, *RGS2*, and *ST3GAL1* all show an increase of at least two fold between the treated and non-treated group and so the combination of this with their low B-H adjusted *p*-values marks them out as potential genes for further investigations. Even larger fold changes are observed as *QRFPR*, *SCN9A*, and *INA* showed over a five-, seven-, and ten-fold increase in expression in platinum-treated patients respectively. As well as genes with substantial increases in gene expression when contrasting treated with untreated patients, there are several genes that show the inverse pattern, i.e. a substantial decrease in those treated with platinum. *GAL*, *UTY*, *AFAP1*, *CYorf15A*, *KDM5D*, *MGC70870*, *RPS4Y1*, *SLC46A3*, and *USP9Y* all show at least a two-fold decrease in the platinum-treated samples when compared with the untreated patients and *RPS4Y1* showed over a six-fold decrease. With a large proportion of these 21 genes showing such large fold changes, their potential importance and changes in their related signalling pathways and networks of genes that they may interact with cannot be overlooked.

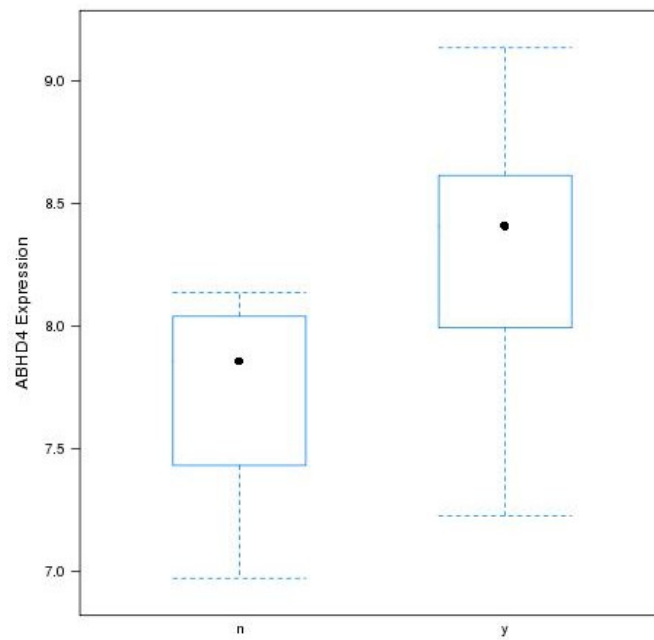
The large fold increase in *INA* expression in patients is corroborated by its appearance in IC50 dose 3 week recovery time course blue WGCNA module, which shows positive correlation with increasing IC50 value (Figure 6.7). *RAB15* and *ABHD4* are also found in this module, as well as the increased expression over the course of the IC50 dose 3 week recovery time course, which is replicated in these patient samples (Figures 6.8 and 6.9).



**Figure 6.7** Box and whisker plot showing the expression of *INA* across the samples, separated into non-platinum-treated (left – n, N=8) and platinum-treated (right – y, N=5). A substantial increase in expression can be seen in the treated samples compared with the untreated.

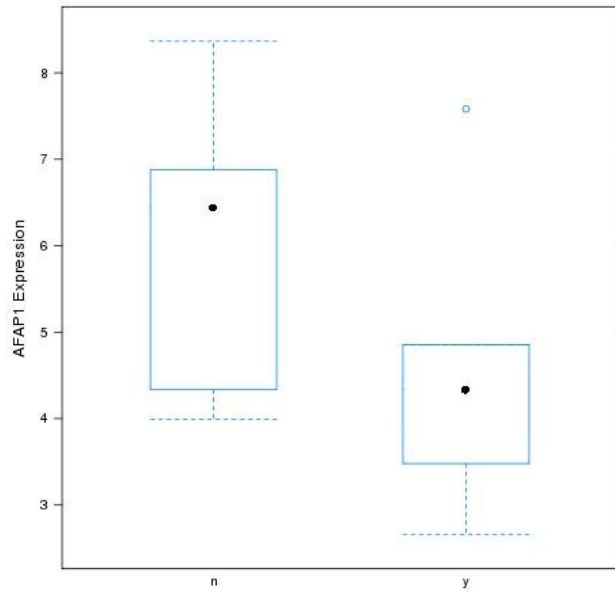


**Figure 6.8** Box and whisker plot showing the expression of *RAB15* across the samples, separated into non-platinum-treated (left – n, N=8) and platinum-treated (right – y, N=5). A substantial increase in expression can be seen in the treated samples compared with the untreated, so much so that the inter-quartile ranges of the two groups do not overlap.



**Figure 6.9** Box and whisker plot showing the expression of *ABHD4* across the samples, separated into non-platinum-treated (left – n, N=8) and platinum-treated (right – y, N=5). A substantial increase in expression can be seen in the treated samples compared with the untreated.

The direction of association with platinum treatment between patient gene expression and the *in vitro* model's gene lists are not always correlated. *AFAP1* lies in the same blue WGCNA module from the IC50 dose 3 week recovery time course as *INA*, *RAB15* and *ABHD4*, however in the patient analyses, a large decrease of over twelve-fold (the reciprocal of 0.0804 Table 6.6) between the treated and untreated patients was observed and can be seen in Figure 6.10.



**Figure 6.10** Box and whisker plot showing the expression of *AFAP1* across the samples, separated into non-platinum-treated (left – n, N=8) and platinum-treated (right – y, N=5) patient samples. A substantial decrease in expression can be seen in the treated samples compared with the untreated.

To investigate this further, the potential hubs (high connectivity genes) from the WGCNA networks (Chapter 5 Section 5.2), were compared with these 21 genes, as well as a further 82 unique transcripts that showed significant at a lower B-H adjusted *p*-value threshold of 0.05 differences in expression between non-platinum treated versus platinum-treated patient samples.

Of the top 21 genes, four appear as hubs in the various modules; *SCN9A*, *AFAP1*, *UTY* and *KDM5D*, two of which (*SCN9A* and *AFAP1*) have very large fold changes when comparing treated and untreated tumour samples. Using the less stringent *p*-value, the number of hubs increases with an additional eight genes (*CGB5*, *CGB*, *GLI2*, *ZNF28*, *DCDC5*, *VSIG1*, *CRISPLD1*, and *ATP1A3*).

When examining these 82 less significant transcripts (B-H adjusted *p*-value of less than 0.05), there are a few particularly interesting candidates. Several genes also fall into functional groups with associations with cancer or signalling pathways that may have effects on the metabolism or transport of drugs or xenobiotics, or that affect the proliferative, transcriptional, and metabolic activities of the tissues.

The known oncogene, *FOS*, is found in the increasing expression cluster of the IC25 dose 1 week recovery time course, and *MXDI*, a histone deacetylase recruiter, may also affect cancer signalling pathways through c-MYC antagonism. *DDIT4* (or *REDD1*) shows a two-fold decrease in platinum-treated samples and is a negative regulator of the mTORC1 complex that in turn is a positive regulator of transcription and protein synthesis. This reduction of *DDIT4* in the presence of platinum could lead to a pro-proliferative signal in these samples and may also allow for cell survival after DNA damage (Vadysirisack *et al.*, 2011).

As mentioned previously Barr *et al.* (2013) have suggested there are alterations in the cell cycle with the accumulation of cisplatin-resistant cells in the G0/G1 phase. The lower significance threshold gene list of this present study shows differential expression of transcripts with potential to influence these processes. *CDKN1A*, found to have nearly a two-fold increase in the treated cohort, is an inhibitor of cyclin-dependent kinases and so is a key regulator of cell cycle progression at G1. It has also been found to be up-regulated in radiotherapy resistant A549 cells (Yang *et al.*, 2013), suggesting this halt at G1 could be key in DNA damage growth arrest in resistant cells. *GADD45B* is also a gene whose transcript levels increase following stressful growth arrest conditions and treatment with DNA-damaging agents, mirrored by its near two-fold increase of the treated patient cohort. This has

also been reported in human colorectal carcinoma (Wang *et al.*, 2012). *RB1*, like *p53*, is vital in control of the cell cycle and it is potentially regulated by another gene in the list of 103 transcripts – *SMYD2*, a lysine methylase. It has a near two-fold decrease in the treated group, which may also be promoting cell cycle arrest by the lack of phosphorylation of *RB1* in the presence of platinum (Cho *et al.*, 2012).

*TMEM45* shows a three-fold increase in treated patients compared to untreated and is thought to be essential for hypoxia-induced chemo-resistance in breast and liver cancer cells (Flamant *et al.*, 2012). *SIDT1*, over two-fold increase in platinum-treated samples, facilitates small RNA transfer between human cells and thus can allow intercellular transfer of a miRNA that drives chemo-resistance (Elhassan *et al.*, 2012).

When the top 21 genes (B-H adjusted *p*-value of less than 0.01) are focused on, similar gene groups can be found. *MAP2K3* and *RAB15* are known to be associated with or involved in cancer-related pathways. The active form of *MAP2K3* is thought to accumulate on expression of the oncogene *KRAS*, and confers oncogenic transformation of primary cells. The *MAPK* family have also been implicated in cellular response to cisplatin and subsequent resistance (Brozovic and Osmak, 2007) and NSCLC cisplatin resistant cell lines have been shown to have higher base line levels of *MAP2K3* protein and potentially inhibit cell death, via p38 *MAPK* on exposure to cisplatin (Galan-Moya *et al.*, 2011). *RAB15* is a member of the same oncogene family and its alternative splicing has been suggested to serve as a biomarker to discriminate tumour-initiating cells from non-tumour-initiating cells in neuroblastoma (Pham *et al.*, 2012).

Signalling pathways and molecules have already been shown to be important in the cellular model and this theme continues with the top 21 genes. *GAL* is a small signalling molecule in

the central and peripheral nervous system (<http://www.genecards.org>) and *QRFPR* is thought to have G-proteins coupled receptor activity and have a role in control of neuroendocrine functions and adipogenesis (Mulumba *et al.*, 2010, Chartrel *et al.*, 2011). *RGS2* as its name suggests, is also involved in G-protein signalling and acts as a GTPase activating protein, driving G proteins into their inactive GDP-bound forms. It has been shown to be down-regulated in the majority of solid tumours including lung cancers (Sethakorn and Dulin, 2013). The PI3K pathway that is associated with proliferation also appears to be altered.

The potential for the increase in cellular activity may also lead to further cellular synthesis, possibly through *C3orf57* and *CYP2SI*, both of which are involved in various lipid synthesis and an increase in the latter is thought to have effects on drug metabolism (Thum *et al.*, 2006).

As well as leading to changes in the DNA damage response and cell cycle, platinum resistance may be the result of alterations in the transport properties of the cells, something that is suggested by the large increase in *SCN9A* expression in the treated samples. Overexpression of *SCN9A* mRNA has already been reported in cervical cancer (Hernandez-Plata *et al.*, 2012). The presence of another relatively unknown solute carrier in top 21 genes (*SLC46A3*) suggests there may be other mechanisms at work that also result in alteration of the potential across membranes. This may be reinforced by the presence of other genes in the lower significance list that have association with chemo-resistance.

The presence of *UTY* and *KDM5D* may indicate differences in epigenetic regulation of the genomics between the treated and untreated samples. Both are histone demethylases, and *KDM5D* has lysine-specific activity and have almost five-fold decreases in expression in the platinum-treated cohort in comparison with the untreated. This is suggestive of a reduction of

histone de-methylation, and in turn of an alteration in chromatin structure and potentially subsequent global gene expression changes. Both are apparent hubs of connectivity (Chapter 5 Section 5.2), which is unsurprising as a change in histone methylation could lead to and correlate with differential expression of several other genes.

*AFAP1*, identified as a hub in the WGCNA analysis, is a binding partner for the tyrosine kinase proto-oncogene *SRC*. It may modulate the actin filament integrity in response to cellular signals and has been suggested to play a role in development and progression of prostate adenocarcinoma by regulating cell-matrix adhesion and migration in the cancer cells (Zhang *et al.*, 2007). The large over twelve-fold decrease seen of this mRNA transcript suggests a substantial change between the treated and untreated cohorts, and potentially the loss of cell-cell interactions or adhesions which may have implications for progressive disease and possibly be indicative of forthcoming move to an invasive phenotype.

It would also be worth investigating these top 21 genes in more depth at both a sequence and epigenetic level to establish what is causing the differences at transcriptome level, such as whether there is DNA variation in the promoter affecting the binding of key transcription factors or promoter methylation. Studies into the downstream effects of the differential expression at the mRNA level would also be of potential interest.

## 6.7 Discussion

This preliminary study looking into a small number of patient samples in an attempt to find validation for results from the previously described cisplatin-resistance cellular model has been relatively successful, in spite of its limitations. A small cohort has been produced for the comparison between platinum-treated and non-platinum-treated tumours of squamous origin.

The small numbers of samples remaining after stringent outlier detection and removal (Section 6.5.1) means the statistical strength of the differential expression investigations performed here is limited. Some genes that do not reach the adjusted statistical significance threshold applied here could nonetheless be involved in the mechanism of response and resistance towards cisplatin and platinum compounds in general, but without more patient samples no firm conclusions can be drawn. Ideally, a specific tissue collection should be set up for patients with NSCLC for both those treated with cisplatin and other platinum treatment regimes, possibly several cycles of treatment, and those of the same age, gender and smoking status that had never been exposed to any platinum or DNA-damaging therapy. This would provide a more suitable cohort of samples for comparison with the cellular models of cisplatin resistance. These types of studies are beginning to be established, but primary care of patients is obviously of primary importance and therefore it is not always feasible to get patients before their first chemotherapy exposure, especially due to the risks associated with lung biopsy.

Despite this present study's limitations, genes of interest from the cell model have been found to have significant and large fold changes of differential expression between the treated and untreated samples. These are found in both the temporal expression clusters and the WGCNA modules of interest that show both positive and negative correlations with the relevant IC

value. Several of these genes appear to have links with various known cancer pathways already. *INA*, which showed a ten-fold increase in platinum-treated samples, has been suggested as a prognostic marker for glial tumours when combined with *p53* and *IDH1* mutations (Takano *et al.*, 2012). *RAB15* is a member of the *RAS* oncogene family that has been implicated in neuroblastoma initiation (Nishimura *et al.*, 2011) and shows two-fold increase in treated patients. The *MAPK* family have already been associated with cisplatin resistance (Brozovic and Osmak, 2007). It could be suggested that others such as *SCN9A* and *SLC46A3*, may be involved in the transport in the cells and therefore alter their phenotype or even affect resistance to cisplatin. *CYP2S1* has also been implicated in xenobiotic metabolism that may affect the cellular response to cisplatin when differentially expressed. It would therefore be worthwhile to explore these 21 genes further in larger cohorts of patients and validate that the changes seen in gene expression occur at the protein level as well.

The substantial fold changes seen for some of the top 21 genes (up to a ten-fold increase or a twelve-fold reduction) combined with their importance shown in network analyses (Chapter 5 Section 5.2) suggest these highly connected genes could be having considerable effects on their environment and influencing other members in their signalling pathways. One could infer these alterations in a handful of genes might be having a large impact on the transcriptome as whole and consequently the tumour on exposure to platinum treatment. It is interesting to note that some of the genes identified as WGCNA hubs (Chapter 5 Section 5.2 and Section 6.6) are present in both the top 21 genes as well as those of lower significance thresholds (B-H adjusted *p*-value of less than 0.05, e.g. *ATPIA3*, *CRISPLD1*, *GLI2*, and *VSIG1*) and may therefore be worthy of further investigation. *AFAP1*, a hub in the top 21 genes with a greater than twelve-fold decrease in treated patients, is thought to act as an adaptor for *SRC* family members, which themselves are implicated in the activation of human

cancer (Irby and Yeatman, 2000). This preliminary study of patient samples, despite its limitations has nevertheless generated some genes worthy of further investigation.

## **7 Discussion and Future Work**

Cancer is a highly diverse disease characterised by multiple mutations that produce the resulting final phenotype seen in an individual patient. Within lung cancer itself there are many different subtypes which may exhibit great heterogeneity in different patients, and in different sites in the same patient and in their subsequent response to treatment. It is therefore highly possible that two individuals, or even two sites of a tumour in the same individual, may not react to the same treatments in the same manner.

In this research project, an attempt has been made to begin to dissect out the complex genomic (RNA expression) response to cisplatin treatment in order to begin to understand the genes and signalling pathways and involved in cisplatin resistance.

### **7.1 Discussion**

#### **7.1.1 A549 Cellular Models of Cisplatin Resistance**

Following extensive optimisation, two robust cellular models of repetitive cisplatin dosing were successfully established using the A549 cell line (Chapter 3). These two models were run over the course of several weeks. At each dosing time point, a measure of resistance (IC<sub>25</sub> or IC<sub>50</sub> value) was calculated and DNA and RNA were harvested from the cells in for in-depth genomic investigation. By synchronisation of cell harvesting and dosing, an overview of the cells' genomic environment at the time of treatment was feasible.

These datasets have yielded detailed information about the population genomic dynamics of the model. Because the response to cisplatin may be, to some extent, random and

heterogeneous the model may reflect the heterogeneity within an individual tumour that is often seen in patient histological specimens.

### 7.1.2 Temporal Gene Expression Profiling

The unique approach of studying profiles of differentially expressed genes over the two treated cisplatin time courses as well as the untreated normal time course, has allowed investigations of changes in global gene expression that may result from cisplatin dosing. The study of multiple time points allowed the evolution of changes within the treatment regime to be investigated. This provides more information than solely focusing on the comparison between resistant and chemo-naïve cells, as is frequently the case in the published literature.

The application of a highly stringent B-H adjusted  $p$ -value has produced genes of robust significance that show distinct changes over the course of cisplatin treatment (Chapter 4 Section 4.3). A number of changes seen may have potentially large effects. Those changes that are simplest to interpret biologically (i.e. increasing and decreasing gene expression with increasing cisplatin doses) have been focused on with the anticipation of finding genes that may be of particular interest.

It is hoped that the removal of genes in common between treated and untreated cells will reduce or even eliminate any changes in transcription that appear simply as a result of extended cell culture (Chapter 4 Section 4.5.2). It is conceivable however that some of these genes may also be of importance in promoting cisplatin resistance despite their presence in untreated cells. It is possible that extended cell culture is a model for cellular immortalisation, which may show hallmarks of cancer, such as cell cycle progression and replication in spite

of DNA damage. The difference in and absence of certain DAVID GO terms between the untreated and treated time courses (Chapter 4 Section 4.6) however, suggest that the systems in action on exposure to cisplatin are distinct from those for continued cell culture.

One limitation of the data from the normal untreated time course is the lack of information on the change in IC<sub>25</sub> or IC<sub>50</sub> value over time. It is possible that the IC value will increase over the repeated rounds of extended cellular passage, irrespective of treatment with cisplatin. If the cells accumulate mutations over this period, it is possible that these changes will alter the response to cisplatin, without being exposed to it, and thus affect their IC value.

The application of highly stringent *p*-values may have had their own limitations in the gene expression analyses. Although the ~33,000 transcripts interrogated by the Affymetrix Gene Arrays needed to be reduced to a manageable number for statistical investigations, it is possible that some low penetrance genetic effects could be masked when using such a harsh statistical threshold. A compromise was made to give a balance between stringency and workability and, several potential genes of biological interest have been found. These include genes encoding the transporter proteins ABCG2, CSAG2, MAL2 and UGT2B15, the adhesion molecules CDH1 and EPCAM, and known cancer genes *EGFR* and *EML4*, all of which have been previously implicated in other cancers. Others have already been shown to be involved in chemo-resistance (the genes *ABCG2*, *CSAG2* and *SFRP1*).

The direct comparisons between the two time courses, with similar numbers of significant differentially expressed genes using the stringent threshold, allowed overlapping genes (N=375) between the two treatment time courses to be identified. These overlapping genes, also not present in the untreated time course, may be the key to the potential mechanisms of survival in the presence of repeated cisplatin dose, irrespective of treatment regime (Chapter

4 Section 4.5.2). The 375 genes were a varied mixture of known cancer genes, transporter proteins, adhesion molecules, and cell cycle proteins. This assortment of genes and the potential subsequently activated pathways gives an indication as to how complex the system of chemo-resistance and continued proliferation and survival is. Nonetheless the 375 genes may prove to be key signalling hubs and therefore are of significant interest.

### 7.1.3 Network Analysis

The use of correlation matrices of gene expression data from the Affymetrix Gene Arrays and the resulting WGCNA have enabled a more general overview of significant differential expression within the cellular models to be generated. Using this approach has meant that the possible interactions between genes were investigated as a whole.

Investigating the correlations in expression profiles via the intermediary of eigengenes allows their (and their corresponding modules) to be compared with the trait of interest which was IC value in this thesis. Consequently it becomes possible to identify hub genes within modules highly positively and negatively correlated with the trait. The hub genes can then be teased out of the large dataset both visually through Circos plots and statistically by investigating their module membership, to give possible candidate genes of interest to be investigated biologically and be compared with the results from the temporal expression clustering.

Four modules were found to be significantly correlated with IC<sub>50</sub> value for the IC<sub>50</sub> dose 3 week recovery time course while the smaller modules for the IC<sub>25</sub> dose 1 week recovery

time course led to correlating modules containing fewer genes and a further reduction of genes of interest (Chapter 5 Sections 5.2.3 and 5.2.4).

Genes already investigated with increasing or decreasing expression over the time course that were found in modules with positive or negative correlation to IC value included the transporters *ABCG2*, *CSAG2* and *UGT2B15* in the IC50 dose 3 week recovery regime. In the IC25 dose 1 week recovery time course, one module contained many genes known to be associated with cancer pathways, including *EML4*. Only *EML4* appeared to be a highly connected hub gene. Other hub genes may not be in the increasing or decreasing temporal expression clusters, but might be exerting their effects on signalling cascades in other ways. For example, other hubs across the treated time courses include the genes *TP63*, *ETS2* and *FOSL2* (Chapter 5, Table 5.5 and Figure 5.14), all of which could be having significant effects on cancer-related pathways to enhance the proliferative and/or cisplatin-resistance phenotype.

The 375 genes overlapping the two treated but not the untreated time courses also underwent network analyses and yielded hub genes of their own including *SCN9A* (another hub gene from the individual time courses) and the gene encoding FGFR1, a receptor known to be involved in cancer-related signalling cascades.

Between the WGCNA analyses and the temporal expression clustering, several genes worthy of further investigation in a more advanced setting were found. A combination of these gene lists of WGCNA modules and temporal expression clusters were therefore the focus when looking at the global gene expression data generated from the patient tissues (Chapter 6 Section 6.6).

#### 7.1.4 Potential Future Work: Cellular Model

##### *7.1.4.1 Other Cellular Models*

The cellular model could be expanded by using other NSCLC cell lines including those with different baseline genetic backgrounds, such as an *EGFR* baseline mutation or one with as few known cancer mutations as possible with the aim of reflecting early stage disease. By expanding the same robust model into different cell line genotypes, a more comprehensive overview of lung cancer responses to cisplatin could be produced. Further comparisons with patients could be performed, such as matching the original subtype of the cell line with tumours of that particular subtype for further tissue-based investigations.

Further studies of the cellular models described in this thesis could also be conducted. RNA extractions after cisplatin dosing and removal could be performed at varying intervals, such as one, four, eight, twelve, sixteen, twenty-four, and forty-eight hours to examine the changing genomic landscape immediately after treatment of chemo-naïve cells and those towards or at the end of the repeated dosing regime. In order to do this, a smaller or shorter cisplatin dose would be required and further optimisation needed to ensure retention of sufficient healthy cells to remain for harvesting and nucleic acid extraction. With the new technology of single cell sequencing of the transcriptome, the limitation of cell numbers is becoming a reality of the future, although this was not feasible during the time of this thesis.

##### *7.1.4.2 Gene Expression over the Time Course*

In this thesis, only fold change differences in gene expression between the start and end points of the time course have been studied in detail. Additional research into the fold

changes of the genes of particular interest across the whole time course could add extra information to the dataset by comparing with chemo-naïve cells and at other points along the dosing regime. These further comparisons across the time course could become highly complex to interpret biologically but may be important; for example, to look into the apparent sigmoidal shape observed around the middle of increasing and decreasing temporal expression clusters in both time courses (Chapter 4, Figures 4.7 and 4.8).

#### *7.1.4.3 Genome of the Cisplatin Resistance Model*

One obvious extension to this work is the need for DNA investigations to combine with the extensive RNA work performed here. As previously stated, DNA was extracted at the same time as RNA from the cells for each dosing time point. DNA aliquots from the start, middle, and end of each of the two time courses have been sent to collaborators with the plan to conduct whole genome sequencing. This would allow structural and sequence changes occurring over the course of treatment to be identified. Examination of whether changes found relate to changes seen in the transcriptome would also be feasible. Results of DNA sequencing for the middle time point could also potentially shed light onto the sigmoidal change observed in the temporal expression clusters and whether there is a genome-related reason behind it.

Ideally these DNA studies would be completed across all the time points of both regimes, as this would enable as full a picture of the genomic landscape as possible to be gained. If RNAseq could also be performed, the entire coding region could be subject to investigation, including non-coding and novel transcripts as well as those that are well documented.

Sequencing of the DNA might also give an indication as to changes in the telomere length over time. It would be interesting to see if the telomeres continued to retract as the model progressed. If this were not the case, one could suggest that the cisplatin treatment was increasing the replicative potential of the cells beyond the normal immortalisation of cell lines.

Other important mechanisms that could be involved in cisplatin resistance involve epigenetics, such as CpG islands or histone modification, which are key in controlling gene expression. A logical next step after conducting Whole Genome Sequencing of the DNA therefore would be to look into global methylation.

#### 7.1.5 Investigating Genes of Interest from the Cellular Model of Resistance between Platinum Treated and Untreated Patients

Two cohorts of cases were included in a preliminary study investigating the differential expression of the genes of interest between patients treated with platinum and those without exposure to any platinum compounds. One cohort was from RESTORE-AIR, a historical NSCLC stent-intervention study, whilst the other was from EQUALITY, an on-going broad spectrum recruitment into many types of lung cancer that is collecting tissue, urine, and blood samples to explore eQTLs. Both RESTORE-AIR and EQUALITY had collected (and continue to collect) tissue samples taken at rigid bronchoscopy with immediate storage of them in RNA stabilisation buffer. This meant that the tissue samples were ideal substrates for transcriptome analyses and would allow the opportunity for comparison of the data generated with the data from the cisplatin-resistance models. For a control normal tissue cohort data generated previously within the group from a project investigating the differentiation between

adenocarcinoma and squamous cell subtypes based on transcriptome signatures/expression profiles was utilised.

Controls were selected based on age, gender, smoking status (current, ex, or never), and subtype to match as closely as possible the RESTORE-AIR and EQUALITY samples. Stringent quality control was performed through methodological outlier detection using the IAC methodology (Chapter 6 Section 6.5.1). The substantial batch effect caused by the marked variances of the studies was removed before differential expression analysis between the platinum-treated and the untreated patients was performed. Unlike the cellular model, a less stringent B-H adjusted  $p$ -value was used to reduce the number of statistically significant differentially expressed genes from the gene lists of interest. By using the gene lists generated by temporal expression clustering and WGCNA (Chapters 4 and 5), the number of comparisons performed in the differential expression analysis was reduced and so a smaller  $p$ -value was applied. Twenty one genes of interest that overlap this patient study and the cellular model were identified. These genes showed highly significant and large fold changes and in several cases have previously been linked with various cancers. Importantly the WGCNA explorations identified a number of them to be key hubs and these genes are therefore potentially good biological candidates for driver mutations in the response against cisplatin.

#### *7.1.5.1 Future Work: Tumour Tissue Samples*

In an ideal world, a specific tissue collection in advanced stage NSCLC patients with 1) a fresh biopsy prior to starting platinum-therapy, 2) a fresh biopsy at time of first response-assessment CT scan, 3) a biopsy at time of completing planned number of platinum

chemotherapy cycles, and 4) with a final fresh biopsy at point of disease relapse would provide substantial insight into and more robust research to compare with the A549 cellular model discussed in previous chapters. This level of interventional biopsy work would however at present not be clinically feasible. Nevertheless, a more limited protocol of biopsy prior to therapy and then again on relapse is clinically achievable. Even with this potentially ideal dataset, there may still be issues with the heterogeneity of cancer as a whole with different patients responding to treatment in variable ways. Patients would also need to be well-phenotyped with substantial follow-up in order to exploit the full potential of such a resource. Additional biopsies of multiple metastatic sites would add further information on potentially relevant resistance mechanisms.

Snap-frozen tissue from these patients and the control cohort, excised at the same time as that used for RNA extraction, is available for DNA extraction. Investigations into differences in structural variation of the DNA of these two cohorts would be worthwhile. If the potential future work discussed in Section 7.1.4.3 was performed using the DNAs from the model, the same investigations could be carried out to provide another comparative cohort to examine whether the DNA variants and changes seen over the course of the cellular model are reflected in the patient samples. These would add to the understanding of why cisplatin resistance is occurring.

Further analyses that could be of use would include examining the differential expression between treated and untreated patients, irrespective of the genes of interest from the cisplatin cellular model, especially in light of the fact that patients may have been exposed to combination therapies. Exploration into the changes in gene expression between these two cohorts alone could provide more known cancer markers and potentially other unknown

biological candidates, which could be further pointers to the mechanisms of survival, and even growth, in the presence of cytotoxic platinum compounds.

The availability of some matched tumour-normal samples from both the RESTORE-AIR and EQUALITY studies, which may be either platinum-treated or non-platinum-treated, would allow comparative analyses to be run to see if there are inherent differences in the host genetics that could drive particular mutations or gene expression profiles observed, as well as their response to chemotherapy and so provide useful insights into disease progression.

## **7.2 General Conclusion**

In spite of the limitations of the cellular model, it has yielded candidate genes for cisplatin resistance that show significant changes in patient tissues that have been exposed to platinum treatment.

The strong overlap in the top hits for genes across the two dosing time courses has identified genes that may be core to resistance to cisplatin. In performing gene expression analyses on A549 cells over extended cell culture, the changes that occur over this period have been controlled for. Variation in the gene lists of differentially expressed genes between the two dosing time courses, as well as some of their patterns of expression in clusters shows there is an individual and stochastic element to the response to cisplatin. Characterisation of the up- and down-regulated genes has been possible by GO analysis and so general mechanisms through which the cells gain resistance have been identified. Network analysis has allowed the differentially expressed genes to be organised and enabled hubs to be found. These genes and hubs may be of therapeutic value.

When examining the patient tissue samples, several genes from the cellular model have been observed that are highly differentially expressed in cisplatin treated squamous cell carcinoma and, as such these genes are potential therapeutic targets. The twenty one genes (discussed in detail in Chapter 6 Section 6.6) are worthy of further investigations, both in an *in vitro* setting, including conducting protein-based assays to assess if the changes at mRNA level are being retained post translation. Further genome-based investigations would also be worthwhile (Section 7.1).

The published research conducted by Barr *et al.* (2013), is to some extent complimentary to investigations performed here. Barr *et al.* (2013) also generated cisplatin-resistant cell lines over a long time period using an isogenic model, but they then characterised the cells using cellular biochemistry techniques as opposed to genetic and genomic approaches. Their comparisons were performed between parent chemo-naïve cells and resistant cells that had been developed over time consistently in media containing cisplatin. They found the cells to have decreased proliferation capacity and as well as observing an accumulation of resistant cells in the G0/G1 phase of the cell cycle. Support of this latter observation comes from the RNA work conducted in this current thesis. In the cellular models the increase in expression of related genes involved in the  $\beta$ -catenin/WNT pathway and the changing abundance of adhesion molecules, reflects the apparent up-regulation of markers for EMT, c-Met, and  $\beta$ -catenin seen by Barr *et al.* (2013). The observation by Barr *et al.* (2013) that the resistant cells had a reduction in cisplatin uptake is potentially supported by this thesis as alterations in the gene expression of transporters were seen in the cellular model, particularly those which have already been implicated in other cancers.

A large database of gene expression data has been generated as a result of this work and it is possible that the highly stringent  $p$ -values implemented (for the cell model data for expression clustering and WGCNA analyses) may have removed low level key players in the survival phenotype. It is hoped that the methodical reasoning applied here has generated results that may help understand the complexity of response to platinum treatment and that the top hits from the large quantity of data produced may lead not only to a better understanding of the mechanisms of cisplatin resistance but provide biomarkers that can be used to help in the prognosis of those going undergoing therapy.

## 8 Bibliography

- ADAMS, V. & HARVEY, R. 2010. Histological and genetic markers for non-small-cell lung cancer: customizing treatment based on individual tumor biology. *Am J Health Syst Pharm*, 67, S3-9, quiz S15-6.
- AHMAD, M., TORKY, A., GLAHN, F., SCHEUBEL, R. & FOTH, H. 2010. PARP-1 expression and activity in primary human lung cells. *Arch Toxicol*.
- AL-GHAMDI, S., ALBASRI, A., CACHAT, J., IBRAHEM, S., MUHAMMAD, B. A., JACKSON, D., NATERI, A. S., KINDLE, K. B. & ILYAS, M. 2011. Cten is targeted by Kras signalling to regulate cell motility in the colon and pancreas. *PLoS One*, 6, e20919.
- AMATO, R., SCUMACI, D., D'ANTONA, L., IULIANO, R., MENNITI, M., DI SANZO, M., FANIELLO, M. C., COLAO, E., MALATESTA, P., ZINGONE, A., AGOSTI, V., COSTANZO, F. S., MILEO, A. M., PAGGI, M. G., LANG, F., CUDA, G., LAVIA, P. & PERROTTI, N. 2012. Sgk1 enhances RANBP1 transcript levels and decreases taxol sensitivity in RKO colon carcinoma cells. *Oncogene*.
- AMOS, C., WU, X., BRODERICK, P., GORLOV, I., GU, J., EISEN, T., DONG, Q., ZHANG, Q., GU, X., VIJAYAKRISHNAN, J., SULLIVAN, K., MATAKIDOU, A., WANG, Y., MILLS, G., DOHENY, K., TSAI, Y., CHEN, W., SHETE, S., SPITZ, M. & HOULSTON, R. 2008. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat Genet*, 40, 616-22.
- ANTONICELLI, A., CAFAROTTI, S., INDINI, A., GALLI, A., RUSSO, A., CESARIO, A., LOCOCO, F. M., RUSSO, P., MAININI, A. F., BONIFATI, L. G., NOSOTTI, M., SANTAMBROGIO, L., MARGARITORA, S., GRANONE, P. M. & DUTLY, A. E. 2013. EGFR-targeted therapy for non-small cell lung cancer: focus on EGFR oncogenic mutation. *Int J Med Sci*, 10, 320-30.
- ARORA, S., KOTHANDAPANI, A., TILLISON, K., KALMAN-MALTESE, V. & PATRICK, S. 2010. Downregulation of XPF-ERCC1 enhances cisplatin efficacy in cancer cells. *DNA Repair (Amst)*, 9, 745-53.
- ASHWORTH, A. 2008. A synthetic lethal therapeutic approach: poly(ADP) ribose polymerase inhibitors for the treatment of cancers deficient in DNA double-strand break repair. *J Clin Oncol*, 26, 3785-90.
- BAI, J. & HU, S. 2012. Transcriptome network analysis reveals potential candidate genes for squamous lung cancer. *Int J Mol Med*, 29, 95-101.
- BANERJEE, P., BASU, A., WEGIEL, B., OTTERBEIN, L. E., MIZUMURA, K., GASSER, M., WAAGA-GASSER, A. M., CHOI, A. M. & PAL, S. 2012. Heme oxygenase-1 promotes survival of renal cancer cells through modulation of apoptosis- and autophagy-regulating molecules. *J Biol Chem*, 287, 32113-23.
- BANKOVIC, J., STOJSIC, J., JOVANOVIĆ, D., ANDJELKOVIC, T., MILINKOVIC, V., RUZDIJIC, S. & TANIC, N. 2010. Identification of genes associated with non-small-cell lung cancer promotion and progression. *Lung Cancer*, 67, 151-9.
- BARR, M. P., GRAY, S. G., HOFFMANN, A. C., HILGER, R. A., THOMALE, J., O'FLAHERTY, J. D., FENNEL, D. A., RICHARD, D., O'LEARY, J. J. & O'BYRNE, K. J. 2013. Generation and characterisation of Cisplatin-resistant non-small cell lung cancer cell lines displaying a stem-like signature. *PLoS One*, 8, e54193.
- BARR, M. P. & O'BYRNE, K. J. 2010. The Generation and Characterisation of Cisplatin-Resistant Non-Small Cell Lung Cancer Cell Lines. *Lung Cancer*, 67.
- BELL, D., GORE, I., OKIMOTO, R., GODIN-HEYMAN, N., SORDELLA, R., MULLOY, R., SHARMA, S., BRANNIGAN, B., MOHAPATRA, G., SETTLEMAN, J. & HABER, D. 2005. Inherited susceptibility to lung cancer may be associated with the T790M drug resistance mutation in EGFR. *Nat Genet*, 37, 1315-6.

- BESSE, B., OLAUSSEN, K. A. & SORIA, J. C. 2013. ERCC1 and RRM1: ready for prime time? *J Clin Oncol*, 31, 1050-60.
- BLUM, C., GRAHAM, A., YOUSEFZADEH, M., SHROUT, J., BENJAMIN, K., KRISHNA, M., HODA, R., COLE, D. J., GARRETT-MAYER, E., REED, C., WALLACE, M. & MITAS, M. 2008. The expression ratio of Map7/B2M is prognostic for survival in patients with stage II colon cancer. *Int J Oncol*, 33, 579-84.
- BROZOVIC, A. & OSMACK, M. 2007. Activation of mitogen-activated protein kinases by cisplatin and their role in cisplatin-resistance. *Cancer Lett*, 251, 1-16.
- CHANG, X., MONITTO, C. L., DEMOKAN, S., KIM, M. S., CHANG, S. S., ZHONG, X., CALIFANO, J. A. & SIDRANSKY, D. 2010. Identification of hypermethylated genes associated with cisplatin resistance in human cancers. *Cancer Res*, 70, 2870-9.
- CHARLES, N. J., THOMAS, P. & LANGE, C. A. 2010. Expression of membrane progesterone receptors (mPR/PAQR) in ovarian cancer cells: implications for progesterone-induced signaling events. *Horm Cancer*, 1, 167-76.
- CHARTREL, N., ALONZEAU, J., ALEXANDRE, D., JEANDEL, L., ALVEAR-PEREZ, R., LEPRINCE, J., BOUTIN, J., VAUDRY, H., ANOUAR, Y. & LLORENS-CORTES, C. 2011. The RFamide neuropeptide 26RFa and its role in the control of neuroendocrine functions. *Front Neuroendocrinol*, 32, 387-97.
- CHEN, S., ZHANG, J., WANG, R., LUO, X. & CHEN, H. 2010. The platinum-based treatments for advanced non-small cell lung cancer, is low/negative ERCC1 expression better than high/positive ERCC1 expression? A meta-analysis. *Lung Cancer*, 70, 63-70.
- CHENG, H., LIU, P., WANG, Z. C., ZOU, L., SANTIAGO, S., GARBITT, V., GJOERUP, O. V., IGLEHART, J. D., MIRON, A., RICHARDSON, A. L., HAHN, W. C. & ZHAO, J. J. 2009. SIK1 couples LKB1 to p53-dependent anoikis and suppresses metastasis. *Sci Signal*, 2, ra35.
- CHEUNG, A. K., KO, J. M., LUNG, H. L., CHAN, K. W., STANBRIDGE, E. J., ZABAROVSKY, E., TOKINO, T., KASHIMA, L., SUZUKI, T., KWONG, D. L., CHUA, D., TSAO, S. W. & LUNG, M. L. 2011. Cysteine-rich intestinal protein 2 (CRIP2) acts as a repressor of NF-kappaB-mediated proangiogenic cytokine transcription to suppress tumorigenesis and angiogenesis. *Proc Natl Acad Sci U S A*, 108, 8390-5.
- CHIQUET, B. T., HENRY, R., BURT, A., MULLIKEN, J. B., STAL, S., BLANTON, S. H. & HECHT, J. T. 2011. Nonsyndromic cleft lip and palate: CRISPLD genes and the folate gene pathway connection. *Birth Defects Res A Clin Mol Teratol*, 91, 44-9.
- CHO, H. S., HAYAMI, S., TOYOKAWA, G., MAEJIMA, K., YAMANE, Y., SUZUKI, T., DOHMAE, N., KOGURE, M., KANG, D., NEAL, D. E., PONDER, B. A., YAMAUE, H., NAKAMURA, Y. & HAMAMOTO, R. 2012. RB1 methylation by SMYD2 enhances cell cycle progression through an increase of RB1 phosphorylation. *Neoplasia*, 14, 476-86.
- COGLIANO, V. J., BAAN, R., STRAIF, K., GROSSE, Y., LAUBY-SECRETAN, B., EL GHISSASSI, F., BOUVARD, V., BENBRAHIM-TALLAA, L., GUHA, N., FREEMAN, C., GALICHET, L. & WILD, C. P. 2011. Preventable exposures associated with human cancers. *J Natl Cancer Inst*, 103, 1827-39.
- COTÉ, M. L., LIU, M., BONASSI, S., NERI, M., SCHWARTZ, A. G., CHRISTIANI, D. C., SPITZ, M. R., MUSCAT, J. E., RENNERT, G., ABEN, K. K., ANDREW, A. S., BENCKO, V., BICKEBÖLLER, H., BOFFETTA, P., BRENNAN, P., BRENNER, H., DUELL, E. J., FABIANOVA, E., FIELD, J. K., FORETOVA, L., FRIIS, S., HARRIS, C. C., HOLCATOVA, I., HONG, Y. C., ISLA, D., JANOUT, V., KIEMENEY, L. A., KIYOHARA, C., LAN, Q., LAZARUS, P., LISSOWSKA, J., LE MARCHAND, L., MATES, D., MATSUO, K., MAYORDOMO, J. I., MCLAUGHLIN, J. R., MORGENSTERN, H., MÜELLER, H., ORLOW, I., PARK, B. J., PINCHEV, M., RAJI, O. Y., RENNERT, H. S., RUDNAL, P., SEOW, A., STUCKER, I., SZESZENIA-DABROWSKA, N., DAWN TEARE, M., TJØNNELAN, A., UGOLINI, D., VAN DER HEIJDEN, H. F., WICHMANN, E., WIENCKE, J. K., WOLL, P. J., YANG, P., ZARIDZE, D., ZHANG, Z. F., ETZEL, C. J. & HUNG, R. J. 2012. Increased risk of lung cancer in

- individuals with a family history of the disease: a pooled analysis from the International Lung Cancer Consortium. *Eur J Cancer*, 48, 1957-68.
- CRIJNS, A. P., DE GRAEFF, P., GEERTS, D., TEN HOOR, K. A., HOLLEMA, H., VAN DER SLUIS, T., HOFSTRA, R. M., DE BOCK, G. H., DE JONG, S., VAN DER ZEE, A. G. & DE VRIES, E. G. 2007. MEIS and PBX homeobox proteins in ovarian cancer. *Eur J Cancer*, 43, 2495-505.
- DAVIES, K. D., LE, A. T., THEODORO, M. F., SKOKAN, M. C., AISNER, D. L., BERGE, E. M., TERRACCIANO, L. M., CAPPUZZO, F., INCARBONE, M., RONCALLI, M., ALLOISIO, M., SANTORO, A., CAMIDGE, D. R., VARELLA-GARCIA, M. & DOEBELE, R. C. 2012. Identifying and Targeting ROS1 Gene Fusions in Non-Small Cell Lung Cancer. *Clin Cancer Res*, 18, 4570-9.
- DEMBÉLÉ, D. & KASTNER, P. 2003. Fuzzy C-means method for clustering microarray data. *Bioinformatics*, 19, 973-80.
- DENNIS, G., SHERMAN, B. T., HOSACK, D. A., YANG, J., GAO, W., LANE, H. C. & LEMPICKI, R. A. 2003. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol*, 4, P3.
- DICKINSON, R. E., FEGAN, K. S., REN, X., HILLIER, S. G. & DUNCAN, W. C. 2011. Glucocorticoid regulation of SLIT/ROBO tumour suppressor genes in the ovarian surface epithelium and ovarian cancer cells. *PLoS One*, 6, e27792.
- DITTMER, S., KOVACS, Z., YUAN, S. H., SISZLER, G., KÖGL, M., SUMMER, H., GEERTS, A., GOLZ, S., SHIODA, T. & METHNER, A. 2011. TOX3 is a neuronal survival factor that induces transcription depending on the presence of CITED1 or phosphorylated CREB in the transcriptionally active complex. *J Cell Sci*, 124, 252-60.
- DO, P. M., VARANASI, L., FAN, S., LI, C., KUBACKA, I., NEWMAN, V., CHAUHAN, K., DANIELS, S. R., BOCCETTA, M., GARRETT, M. R., LI, R. & MARTINEZ, L. A. 2012. Mutant p53 cooperates with ETS2 to promote etoposide resistance. *Genes Dev*, 26, 830-45.
- ELHASSAN, M. O., CHRISTIE, J. & DUXBURY, M. S. 2012. Homo sapiens systemic RNA interference-defective-1 transmembrane family member 1 (SIDT1) protein mediates contact-dependent small RNA transfer and microRNA-21-driven chemoresistance. *J Biol Chem*, 287, 5267-77.
- ELZI, D. J., SONG, M., HAKALA, K., WEINTRAUB, S. T. & SHIO, Y. 2012. Wnt antagonist SFRP1 functions as a secreted mediator of senescence. *Mol Cell Biol*, 32, 4388-99.
- ENG, L., IBRAHIM-ZADA, I., JARJANAZI, H., SAVAS, S., MESCHIAN, M., PRITCHARD, K. I. & OZCELIK, H. 2011. Bioinformatic analyses identifies novel protein-coding pharmacogenomic markers associated with paclitaxel sensitivity in NCI60 cancer cell lines. *BMC Med Genomics*, 4, 18.
- ENGELMAN, J., ZEJNULLAHU, K., MITSUDOMI, T., SONG, Y., HYLAND, C., PARK, J., LINDEMAN, N., GALE, C., ZHAO, X., CHRISTENSEN, J., KOSAKA, T., HOLMES, A., ROGERS, A., CAPPUZZO, F., MOK, T., LEE, C., JOHNSON, B., CANTLEY, L. & JÄNNE, P. 2007. MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling. *Science*, 316, 1039-43.
- EYMIN, B. & GAZZERI, S. 2010. Role of cell cycle regulators in lung carcinogenesis. *Cell Adh Migr*, 4, 114-23.
- FENG, J., ZHANG, X., ZHU, H., WANG, X., NI, S. & HUANG, J. 2012a. FoxQ1 overexpression influences poor prognosis in non-small cell lung cancer, associates with the phenomenon of EMT. *PLoS One*, 7, e39937.
- FENG, Y., WANG, Y., WANG, Z., FANG, Z., LI, F., GAO, Y., LIU, H., XIAO, T., ZHOU, Y., ZHAI, Q., LIU, X., SUN, Y., BARDEESY, N., WONG, K. K., CHEN, H., XIONG, Z. Q. & JI, H. 2012b. The CRTC1-NEDD9 signaling axis mediates lung cancer progression caused by LKB1 loss. *Cancer Res*, 72, 6502-11.
- FITZGIBBON, J., SMITH, L. L., RAGHAVAN, M., SMITH, M. L., DEBERNARDI, S., SKOULAKIS, S., LILLINGTON, D., LISTER, T. A. & YOUNG, B. D. 2005. Association

- between acquired uniparental disomy and homozygous gene mutation in acute myeloid leukemias. *Cancer Res*, 65, 9152-4.
- FLAMANT, L., ROEGIERS, E., PIERRE, M., HAYEZ, A., STERPIN, C., DE BACKER, O., ARNOULD, T., POUMAY, Y. & MICHIELS, C. 2012. TMEM45A is essential for hypoxia-induced chemoresistance in breast and liver cancer cells. *BMC Cancer*, 12, 391.
- FLEMING, I. N., ELLIOTT, C. M., COLLARD, J. G. & EXTON, J. H. 1997. Lysophosphatidic acid induces threonine phosphorylation of Tiam1 in Swiss 3T3 fibroblasts via activation of protein kinase C. *J Biol Chem*, 272, 33105-10.
- FRIBOULET, L., OLAUSSEN, K. A., PIGNON, J. P., SHEPHERD, F. A., TSAO, M. S., GRAZIANO, S., KRATZKE, R., DOUILLARD, J. Y., SEYMOUR, L., PIRKER, R., FILIPITS, M., ANDRÉ, F., SOLARY, E., PONSONNAILLES, F., ROBIN, A., STOCLIN, A., DORVAULT, N., COMMO, F., ADAM, J., VANHECKE, E., SAULNIER, P., THOMALE, J., LE CHEVALIER, T., DUNANT, A., ROUSSEAU, V., LE TEUFF, G., BRAMBILLA, E. & SORIA, J. C. 2013. ERCC1 isoform expression and DNA repair in non-small-cell lung cancer. *N Engl J Med*, 368, 1101-10.
- FRIEDBERG, E. C. 2001. How nucleotide excision repair protects against cancer. *Nat Rev Cancer*, 1, 22-33.
- FRYER, R. M., RANDALL, J., YOSHIDA, T., HSIAO, L. L., BLUMENSTOCK, J., JENSEN, K. E., DIMOFTE, T., JENSEN, R. V. & GULLANS, S. R. 2002. Global analysis of gene expression: methods, interpretation, and pitfalls. *Exp Nephrol*, 10, 64-74.
- FU, J., FONG, K., BELLACOSA, A., ROSS, E., APOSTOLOU, S., BASSI, D. E., JIN, F., ZHANG, J., CAIRNS, P., IBAÑEZ DE CACERES, I., BRAUNEWELL, K. H. & KLEIN-SZANTO, A. J. 2008. VILIP-1 downregulation in non-small cell lung carcinomas: mechanisms and prediction of survival. *PLoS One*, 3, e1698.
- GALAN-MOYA, E. M., DE LA CRUZ-MORCILLO, M. A., LLANOS VALERO, M., CALLEJAS-VALERA, J. L., MELGAR-ROJAS, P., HERNADEZ LOSA, J., SALCEDO, M., FERNÁNDEZ-ARAMBURO, A., RAMON Y CAJAL, S. & SÁNCHEZ-PRIETO, R. 2011. Balance between MKK6 and MKK3 mediates p38 MAPK associated resistance to cisplatin in NSCLC. *PLoS One*, 6, e28406.
- GALLUZZI, L., MARSILI, S., VITALE, I., SENOVILLA, L., MICHELS, J., GARCIA, P., VACCHELLI, E., CHATELUT, E., CASTEDO, M. & KROEMER, G. 2013a. Vitamin B6 metabolism influences the intracellular accumulation of cisplatin. *Cell Cycle*, 12, 417-21.
- GALLUZZI, L., SENOVILLA, L., VITALE, I., MICHELS, J., MARTINS, I., KEPP, O., CASTEDO, M. & KROEMER, G. 2012. Molecular mechanisms of cisplatin resistance. *Oncogene*, 31, 1869-83.
- GALLUZZI, L., VACCHELLI, E., MICHELS, J., GARCIA, P., KEPP, O., SENOVILLA, L., VITALE, I. & KROEMER, G. 2013b. Effects of vitamin B6 metabolism on oncogenesis, tumor progression and therapeutic responses. *Oncogene*.
- GANDHI, J., ZHANG, J., XIE, Y., SOH, J., SHIGEMATSU, H., ZHANG, W., YAMAMOTO, H., PEYTON, M., GIRARD, L., LOCKWOOD, W., LAM, W., VARELLA-GARCIA, M., MINNA, J. & GAZDAR, A. 2009. Alterations in genes of the EGFR signaling pathway and their relationship to EGFR tyrosine kinase inhibitor sensitivity in lung cancer cell lines. *PLoS One*, 4, e4576.
- GAZDAR, A. 2007. DNA repair and survival in lung cancer--the two faces of Janus. *N Engl J Med*, 356, 771-3.
- GIARD, D. J., AARONSON, S. A., TODARO, G. J., ARNSTEIN, P., KERSEY, J. H., DOSIK, H. & PARKS, W. P. 1973. In vitro cultivation of human tumors: establishment of cell lines derived from a series of solid tumors. *J Natl Cancer Inst*, 51, 1417-23.
- GILL, R. K., YANG, S. H., MEERZAMAN, D., MECHANIC, L. E., BOWMAN, E. D., JEON, H. S., ROY CHOWDHURI, S., SHAKOORI, A., DRACHEVA, T., HONG, K. M., FUKUOKA, J., ZHANG, J. H., HARRIS, C. C. & JEN, J. 2011. Frequent homozygous deletion of the LKB1/STK11 gene in non-small cell lung cancer. *Oncogene*, 30, 3784-91.

- GOLDSTRAW, P., BALL, D., JETT, J. R., LE CHEVALIER, T., LIM, E., NICHOLSON, A. G. & SHEPHERD, F. A. 2011. Non-small-cell lung cancer. *Lancet*, 378, 1727-40.
- GOLDSTRAW, P., CROWLEY, J., CHANSKY, K., GIROUX, D., GROOME, P., RAMI-PORTA, R., POSTMUS, P., RUSCH, V. & SOBIN, L. 2007. The IASLC Lung Cancer Staging Project: proposals for the revision of the TNM stage groupings in the forthcoming (seventh) edition of the TNM Classification of malignant tumours. *J Thorac Oncol*, 2, 706-14.
- GUPTA, N., MIYAUCHI, S., MARTINDALE, R. G., HERDMAN, A. V., PODOLSKY, R., MIYAKE, K., MAGER, S., PRASAD, P. D., GANAPATHY, M. E. & GANAPATHY, V. 2005. Upregulation of the amino acid transporter ATB<sub>0,+</sub> (SLC6A14) in colorectal cancer and metastasis in humans. *Biochim Biophys Acta*, 1741, 215-23.
- GUPTA, N., PRASAD, P. D., GHAMANDE, S., MOORE-MARTIN, P., HERDMAN, A. V., MARTINDALE, R. G., PODOLSKY, R., MAGER, S., GANAPATHY, M. E. & GANAPATHY, V. 2006. Up-regulation of the amino acid transporter ATB<sub>0,+</sub> (SLC6A14) in carcinoma of the cervix. *Gynecol Oncol*, 100, 8-13.
- HAJJARI, M., BEHMANESH, M., SADEGHIZADEH, M. & ZEINODDINI, M. 2013. Junctional adhesion molecules 2 and 3 may potentially be involved in progression of gastric adenocarcinoma tumors. *Med Oncol*, 30, 380.
- HALL, B. A., KIM, T. Y., SKOR, M. N. & CONZEN, S. D. 2012. Serum and glucocorticoid-regulated kinase 1 (SGK1) activation in breast cancer: requirement for mTORC1 activity associates with ER- $\alpha$  expression. *Breast Cancer Res Treat*, 135, 469-79.
- HAMMERMAN, P. S., SOS, M. L., RAMOS, A. H., XU, C., DUTT, A., ZHOU, W., BRACE, L. E., WOODS, B. A., LIN, W., ZHANG, J., DENG, X., LIM, S. M., HEYNCK, S., PEIFER, M., SIMARD, J. R., LAWRENCE, M. S., ONOFRIO, R. C., SALVESEN, H. B., SEIDEL, D., ZANDER, T., HEUCKMANN, J. M., SOLTERMANN, A., MOCH, H., KOKER, M., LEENDERS, F., GABLER, F., QUERINGS, S., ANSEN, S., BRAMBILLA, E., BRAMBILLA, C., LORIMIER, P., BRUSTUGUN, O. T., HELLAND, A., PETERSEN, I., CLEMENT, J. H., GROEN, H., TIMENS, W., SIETSMA, H., STOELBEN, E., WOLF, J., BEER, D. G., TSAO, M. S., HANNA, M., HATTON, C., ECK, M. J., JANNE, P. A., JOHNSON, B. E., WINCKLER, W., GREULICH, H., BASS, A. J., CHO, J., RAUH, D., GRAY, N. S., WONG, K. K., HAURA, E. B., THOMAS, R. K. & MEYERSON, M. 2011. Mutations in the DDR2 kinase gene identify a novel therapeutic target in squamous cell lung cancer. *Cancer Discov*, 1, 78-89.
- HANAHAN, D. & WEINBERG, R. 2000. The hallmarks of cancer. *Cell*, 100, 57-70.
- HEIN, S., MAHNER, S., KANOWSKI, C., LÖNING, T., JÄNICKE, F. & MILDE-LANGOSCH, K. 2009. Expression of Jun and Fos proteins in ovarian tumors of different malignant potential and in ovarian cancer cell lines. *Oncol Rep*, 22, 177-83.
- HERBST, R., HEYMACH, J. & LIPPMAN, S. 2008. Lung cancer. *N Engl J Med*, 359, 1367-80.
- HERNANDEZ-PLATA, E., ORTIZ, C. S., MARQUINA-CASTILLO, B., MEDINA-MARTINEZ, I., ALFARO, A., BERUMEN, J., RIVERA, M. & GOMORA, J. C. 2012. Overexpression of NaV 1.6 channels is associated with the invasion capacity of human cervical cancer. *Int J Cancer*, 130, 2013-23.
- HSU, D., ACHARYA, C., BALAKUMARAN, B., RIEDEL, R., KIM, M., STEVENSON, M., TUCHMAN, S., MUKHERJEE, S., BARRY, W., DRESSMAN, H., NEVINS, J., POWERS, S., MU, D. & POTTI, A. 2009. Characterizing the developmental pathways TTF-1, NKX2-8, and PAX9 in lung cancer. *Proc Natl Acad Sci U S A*, 106, 5312-7.
- HUANG, Y., ZHANG, J., ZHAO, Y. Y., JIANG, W., XUE, C., XU, F., ZHAO, H. Y., ZHANG, Y., ZHAO, L. P., HU, Z. H., YAO, Z. W., LIU, Q. Y. & ZHANG, L. 2012. SPARC expression and prognostic value in non-small cell lung cancer. *Chin J Cancer*, 31, 541-8.
- HUNG, R., MCKAY, J., GABORIEAU, V., BOFFETTA, P., HASHIBE, M., ZARIDZE, D., MUKERIA, A., SZESZENIA-DABROWSKA, N., LISSOWSKA, J., RUDNAI, P., FABIANOVA, E., MATES, D., BENCKO, V., FORETOVA, L., JANOUT, V., CHEN, C., GOODMAN, G., FIELD, J., LILOGLOU, T., XINARIANOS, G., CASSIDY, A.,

- MCLAUGHLIN, J., LIU, G., NAROD, S., KROKAN, H., SKORPEN, F., ELVESTAD, M., HVEEM, K., VATTEN, L., LINSEISEN, J., CLAVEL-CHAPELON, F., VINEIS, P., BUENO-DE-MESQUITA, H., LUND, E., MARTINEZ, C., BINGHAM, S., RASMUSON, T., HAINAUT, P., RIBOLI, E., AHRENS, W., BENHAMOU, S., LAGIOU, P., TRICHOPOULOS, D., HOLCÁTOVÁ, I., MERLETTI, F., KJAERHEIM, K., AGUDO, A., MACFARLANE, G., TALAMINI, R., SIMONATO, L., LOWRY, R., CONWAY, D., ZNAOR, A., HEALY, C., ZELENKA, D., BOLAND, A., DELEPINE, M., FOGLIO, M., LECHNER, D., MATSUDA, F., BLANCHE, H., GUT, I., HEATH, S., LATHROP, M. & BRENNAN, P. 2008. A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature*, 452, 633-7.
- IKEDIOBI, O., DAVIES, H., BIGNELL, G., EDKINS, S., STEVENS, C., O'MEARA, S., SANTARIUS, T., AVIS, T., BARTHORPE, S., BRACKENBURY, L., BUCK, G., BUTLER, A., CLEMENTS, J., COLE, J., DICKS, E., FORBES, S., GRAY, K., HALLIDAY, K., HARRISON, R., HILLS, K., HINTON, J., HUNTER, C., JENKINSON, A., JONES, D., KOSMIDOU, V., LUGG, R., MENZIES, A., MIRONENKO, T., PARKER, A., PERRY, J., RAINE, K., RICHARDSON, D., SHEPHERD, R., SMALL, A., SMITH, R., SOLOMON, H., STEPHENS, P., TEAGUE, J., TOFTS, C., VARIAN, J., WEBB, T., WEST, S., WIDAA, S., YATES, A., REINHOLD, W., WEINSTEIN, J., STRATTON, M., FUTREAL, P. & WOOSTER, R. 2006. Mutation analysis of 24 known cancer genes in the NCI-60 cell line set. *Mol Cancer Ther*, 5, 2606-12.
- IRBY, R. B. & YEATMAN, T. J. 2000. Role of Src expression and activation in human cancer. *Oncogene*, 19, 5636-42.
- IWASAKI, H., NAKANO, K., SHINKAI, K., KUNISAWA, Y., HIRAHASHI, M., ODA, Y., ONISHI, H. & KATANO, M. 2012. Hedgehog Gli3 activator signal augments tumorigenicity of colorectal cancer via upregulation of adherence-related genes. *Cancer Sci*.
- JAVIDROOZI, M., ZUCKER, S. & CHEN, W. T. 2012. Plasma seprase and DPP4 levels as markers of disease and prognosis in cancer. *Dis Markers*, 32, 309-20.
- JH, H., RAMSEY, M., HAYES, D., FAN, C., MCNAMARA, K., KOZLOWSKI, P., TORRICE, C., WU, M., SHIMAMURA, T., PERERA, S., LIANG, M., CAI, D., NAUMOV, G., BAO, L., CONTRERAS, C., LI, D., CHEN, L., KRISHNAMURTHY, J., KOIVUNEN, J., CHIRIEAC, L., PADERA, R., BRONSON, R., LINDEMAN, N., CHRISTIANI, D., LIN, X., SHAPIRO, G., JÄNNE, P., JOHNSON, B., MEYERSON, M., KWIATKOWSKI, D., CASTRILLON, D., BARDEESY, N., SHARPLESS, N. & WONG, K. 2007. LKB1 modulates lung cancer differentiation and metastasis. *Nature*, 448, 807-10.
- JIA, M., SOUCHELNYTSKYI, N., HELLMAN, U., O'HARE, M., JAT, P. S. & SOUCHELNYTSKYI, S. 2010. Proteome profiling of immortalization-to-senescence transition of human breast epithelial cells identified MAP2K3 as a senescence-promoting protein which is downregulated in human breast cancer. *Proteomics Clin Appl*, 4, 816-28.
- JU, W., YOO, B. C., KIM, I. J., KIM, J. W., KIM, S. C. & LEE, H. P. 2009. Identification of genes with differential expression in chemoresistant epithelial ovarian cancer using high-density oligonucleotide microarrays. *Oncol Res*, 18, 47-56.
- KARANJAWALA, Z. E., ILLEI, P. B., ASHFAQ, R., INFANTE, J. R., MURPHY, K., PANDEY, A., SCHULICK, R., WINTER, J., SHARMA, R., MAITRA, A., GOGGINS, M. & HRUBAN, R. H. 2008. New markers of pancreatic cancer identified through differential gene expression analyses: claudin 18 and annexin A8. *Am J Surg Pathol*, 32, 188-96.
- KARUNAKARAN, S., RAMACHANDRAN, S., COOTHANKANDASWAMY, V., ELANGO VAN, S., BABU, E., PERIYASAMY-THANDAVAN, S., GURAV, A., GNANAPRAKASAM, J. P., SINGH, N., SCHOENLEIN, P. V., PRASAD, P. D., THANGARAJU, M. & GANAPATHY, V. 2011. SLC6A14 (ATB0,+) protein, a highly concentrative and broad specific amino acid transporter, is a novel and effective drug target for treatment of estrogen receptor-positive breast cancer. *J Biol Chem*, 286, 31830-8.

- KNAPPSKOG, S., MYKLEBUST, L. M., BUSCH, C., ALOYSIUS, T., VARHAUG, J. E., LØNNING, P. E., LILLEHAUG, J. R. & PENDINO, F. 2011. RINF (CXXC5) is overexpressed in solid tumors and is an unfavorable prognostic factor in breast cancer. *Ann Oncol*, 22, 2208-15.
- KOHLER, L. H., MIRESKANDARI, M., KNÖSEL, T., ALTENDORF-HOFMANN, A., KUNZE, A., SCHMIDT, A., PRESSELT, N., CHEN, Y. & PETERSEN, I. 2012. FGFR1 expression and gene copy numbers in human lung cancer. *Virchows Arch*, 461, 49-57.
- KOKUBO, Y., GEMMA, A., NORO, R., SEIKE, M., KATAOKA, K., MATSUDA, K., OKANO, T., MINEGISHI, Y., YOSHIMURA, A., SHIBUYA, M. & KUDOH, S. 2005. Reduction of PTEN protein and loss of epidermal growth factor receptor gene mutation in lung cancer with natural resistance to gefitinib (IRESSA). *Br J Cancer*, 92, 1711-9.
- KONDO, S., IWATA, S., YAMADA, T., INOUE, Y., ICHIHARA, H., KICHIKAWA, Y., KATAYOSE, T., SOUTA-KURIBARA, A., YAMAZAKI, H., HOSONO, O., KAWASAKI, H., TANAKA, H., HAYASHI, Y., SAKAMOTO, M., KAMIYA, K., DANG, N. H. & MORIMOTO, C. 2012. Impact of the integrin signaling adaptor protein NEDD9 on prognosis and metastatic behavior of human lung cancer. *Clin Cancer Res*, 18, 6326-38.
- KOTOULA, V., KRIKELIS, D., KARAVASILIS, V., KOLETSA, T., ELEFThERAKI, A. G., TELEVANTOU, D., CHRISTODOULOU, C., DIMOUDIS, S., KORANTZIS, I., PECTASIDES, D., SYRIGOS, K. N., KOSMIDIS, P. A. & FOUNTZILAS, G. 2012. Expression of DNA repair and replication genes in non-small cell lung cancer (NSCLC): a role for thymidylate synthetase (TYMS). *BMC Cancer*, 12, 342.
- KOWANETZ, M., LÖNN, P., VANLANDEWIJCK, M., KOWANETZ, K., HELDIN, C. H. & MOUSTAKAS, A. 2008. TGFbeta induces SIK to negatively regulate type I receptor kinase signaling. *J Cell Biol*, 182, 655-62.
- KUMAR, L. & E FUTSCHIK, M. 2007. Mfuzz: a software package for soft clustering of microarray data. *Bioinformatics*, 2, 5-7.
- KUUSELO, R., SAVINAINEN, K., SANDSTRÖM, S., AUTIO, R. & KALLIONIEMI, A. 2011. MED29, a component of the mediator complex, possesses both oncogenic and tumor suppressive characteristics in pancreatic cancer. *Int J Cancer*, 129, 2553-65.
- LANGFELDER, P. & HORVATH, S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, 9, 559.
- LI, B., WEN, G., ZHAO, Y., TONG, J. & HEI, T. K. 2012. The role of TGFBI in mesothelioma and breast cancer: association with tumor suppression. *BMC Cancer*, 12, 239.
- LIU, C., LI, Q., ZHOU, X., KOLOSOV, V. P. & PERELMAN, J. M. 2013. Regulator of G-protein signaling 2 inhibits acid-induced mucin5AC hypersecretion in human airway epithelial cells. *Respir Physiol Neurobiol*, 185, 265-71.
- LIU, Y., CHEN, J., YANG, Y., ZHANG, L. & JIANG, W. G. 2012. Molecular impact of bone morphogenetic protein 7, on lung cancer cells and its clinical significance. *Int J Mol Med*, 29, 1016-24.
- LOCK, F. E., UNDERHILL-DAY, N., DUNWELL, T., MATAILLANAS, D., COOPER, W., HESSON, L., RECINO, A., WARD, A., PAVLOVA, T., ZABAROVSKY, E., GRANT, M. M., MAHER, E. R., CHALMERS, A. D., KOLCH, W. & LATIF, F. 2010. The RASSF8 candidate tumor suppressor inhibits cell growth and regulates the Wnt and NF-kappaB signaling pathways. *Oncogene*, 29, 4307-16.
- LORIOT, Y., MORDANT, P., DEUTSCH, E., OLAUSSEN, K. & SORIA, J. 2009. Are RAS mutations predictive markers of resistance to standard chemotherapy? *Nat Rev Clin Oncol*, 6, 528-34.
- LU, Y. C., CHEN, C. N., WANG, B., HSU, W. M., CHEN, S. T., CHANG, K. J., CHANG, C. C. & LEE, H. 2011. Changes in tumor growth and metastatic capacities of J82 human bladder cancer cells suppressed by down-regulation of calreticulin expression. *Am J Pathol*, 179, 1425-33.

- LUO, D., LU, M. L., ZHAO, G. F., HUANG, H., ZHENG, M. Y., CHANG, J., LV, L. & LUO, J. B. 2012. Reduced Popdc3 expression correlates with high risk and poor survival in patients with gastric cancer. *World J Gastroenterol*, 18, 2423-9.
- LYNCH, T., BELL, D., SORDELLA, R., GURUBHAGAVATULA, S., OKIMOTO, R., BRANNIGAN, B., HARRIS, P., HASERLAT, S., SUPKO, J., HALUSKA, F., LOUIS, D., CHRISTIANI, D., SETTLEMAN, J. & HABER, D. 2004. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med*, 350, 2129-39.
- MARSHALL, J. C., COLLINS, J. W., NAKAYAMA, J., HORAK, C. E., LIEWEHR, D. J., STEINBERG, S. M., ALBAUGH, M., VIDAL-VANACLOCHA, F., PALMIERI, D., BARBIER, M., MURONE, M. & STEEG, P. S. 2012. Effect of inhibition of the lysophosphatidic acid receptor 1 on metastasis and metastatic dormancy in breast cancer. *J Natl Cancer Inst*, 104, 1306-19.
- MARTOWICZ, A., SPIZZO, G., GASTL, G. & UNTERGASSER, G. 2012. Phenotype-dependent effects of EpCAM expression on growth and invasion of human breast cancer cell lines. *BMC Cancer*, 12, 501.
- MASCAUX, C., IANNINO, N., MARTIN, B., PAESMANS, M., BERGHMANS, T., DUSART, M., HALLER, A., LOTHAIRE, P., MEERT, A., NOEL, S., LAFITTE, J. & SCULIER, J. 2005. The role of RAS oncogene in survival of patients with lung cancer: a systematic review of the literature with meta-analysis. *Br J Cancer*, 92, 131-9.
- MASTERS, J. R. 2000. Human cancer cell lines: fact and fantasy. *Nat Rev Mol Cell Biol*, 1, 233-6.
- MATERNA, V., SUROWIAK, P., KAPLENKO, I., SPACZYŃSKI, M., DUAN, Z., ZABEL, M., DIETEL, M. & LAGE, H. 2007. Taxol-resistance-associated gene-3 (TRAG-3/CSAG2) expression is predictive for clinical outcome in ovarian carcinoma patients. *Virchows Arch*, 450, 187-94.
- MCKAY, J., HUNG, R., GABORIEAU, V., BOFFETTA, P., CHABRIER, A., BYRNES, G., ZARIDZE, D., MUKERIA, A., SZESZENIA-DABROWSKA, N., LISSOWSKA, J., RUDNAI, P., FABIANOVA, E., MATES, D., BENCKO, V., FORETOVA, L., JANOUT, V., MCLAUGHLIN, J., SHEPHERD, F., MONTPETIT, A., NAROD, S., KROKAN, H., SKORPEN, F., ELVESTAD, M., VATTEN, L., NJØLSTAD, I., AXELSSON, T., CHEN, C., GOODMAN, G., BARNETT, M., LOOMIS, M., LUBIŃSKI, J., MATYJASIK, J., LENER, M., OSZUTOWSKA, D., FIELD, J., LILOGLOU, T., XINARIANOS, G., CASSIDY, A., VINEIS, P., CLAVEL-CHAPELON, F., PALLI, D., TUMINO, R., KROGH, V., PANICO, S., GONZÁLEZ, C., RAMÓN QUIRÓS, J., MARTÍNEZ, C., NAVARRO, C., ARDANAZ, E., LARRAÑAGA, N., KHAM, K., KEY, T., BUENO-DE-MESQUITA, H., PEETERS, P., TRICHOPOULOU, A., LINSEISEN, J., BOEING, H., HALLMANS, G., OVERVAD, K., TJØNNELAND, A., KUMLE, M., RIBOLI, E., ZELENKA, D., BOLAND, A., DELEPINE, M., FOGGIO, M., LECHNER, D., MATSUDA, F., BLANCHE, H., GUT, I., HEATH, S., LATHROP, M. & BRENNAN, P. 2008. Lung cancer susceptibility locus at 5p15.33. *Nat Genet*, 40, 1404-6.
- MITSUDOMI, T., VIALLET, J., MULSHINE, J., LINNOILA, R., MINNA, J. & GAZDAR, A. 1991. Mutations of ras genes distinguish a subset of non-small-cell lung cancer cell lines from small-cell lung cancer cell lines. *Oncogene*, 6, 1353-62.
- MULUMBA, M., JOSSART, C., GRANATA, R., GALLO, D., ESCHER, E., GHIGO, E., SERVANT, M. J., MARLEAU, S. & ONG, H. 2010. GPR103b functions in the peripheral regulation of adipogenesis. *Mol Endocrinol*, 24, 1615-25.
- NISHIMURA, N., VAN HUYEN PHAM, T., HARTOMO, T. B., LEE, M. J., HASEGAWA, D., TAKEDA, H., KAWASAKI, K., KOSAKA, Y., YAMAMOTO, T., MORIKAWA, S., YAMAMOTO, N., KUBOKAWA, I., MORI, T., YANAI, T., HAYAKAWA, A., TAKESHIMA, Y., NISHIO, H. & MATSUO, M. 2011. Rab15 expression correlates with retinoic acid-induced differentiation of neuroblastoma cells. *Oncol Rep*, 26, 145-51.

- OKAMURA, J., HUANG, Y., MOON, D., BRAIT, M., CHANG, X. & KIM, M. S. 2012. Downregulation of insulin-like growth factor-binding protein 7 in cisplatin-resistant non-small cell lung cancer. *Cancer Biol Ther*, 13, 148-55.
- OSHLACK, A., ROBINSON, M. D. & YOUNG, M. D. 2010. From RNA-seq reads to differential expression results. *Genome Biol*, 11, 220.
- OXNARD, G. R., BINDER, A. & JÄNNE, P. A. 2013. New Targetable Oncogenes in Non-Small-Cell Lung Cancer. *J Clin Oncol*.
- PAEZ, J., JÄNNE, P., LEE, J., TRACY, S., GREULICH, H., GABRIEL, S., HERMAN, P., KAYE, F., LINDEMAN, N., BOGGON, T., NAOKI, K., SASAKI, H., FUJII, Y., ECK, M., SELLERS, W., JOHNSON, B. & MEYERSON, M. 2004. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science*, 304, 1497-500.
- PAO, W., MILLER, V., POLITI, K., RIELY, G., SOMWAR, R., ZAKOWSKI, M., KRIS, M. & VARMUS, H. 2005a. Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. *PLoS Med*, 2, e73.
- PAO, W., WANG, T., RIELY, G., MILLER, V., PAN, Q., LADANYI, M., ZAKOWSKI, M., HEELAN, R., KRIS, M. & VARMUS, H. 2005b. KRAS mutations and primary resistance of lung adenocarcinomas to gefitinib or erlotinib. *PLoS Med*, 2, e17.
- PARKIN, D. M. 2011. 2. Tobacco-attributable cancer burden in the UK in 2010. *Br J Cancer*, 105 Suppl 2, S6-S13.
- PARON, I., BERCHTOLD, S., VÖRÖS, J., SHAMARLA, M., ERKAN, M., HÖFLER, H. & ESPOSITO, I. 2011. Tenascin-C enhances pancreatic cancer cell growth and motility and affects cell adhesion through activation of the integrin pathway. *PLoS One*, 6, e21684.
- PAUL, I., WANG, C., SAVAGE, K., LAMERS, E., GATELY, K., O'BYRNE, K. J., QUINN, J. & FENNELL, D. A. Synthetic Lethal Targeting of BRCA1-immunodeficient Non-Small Cell Lung Cancer by Poly ADP Ribose Polymerase Inhibition. *Molecular Markers*, 2010.
- PHAM, T. V., HARTOMO, T. B., LEE, M. J., HASEGAWA, D., ISHIDA, T., KAWASAKI, K., KOSAKA, Y., YAMAMOTO, T., MORIKAWA, S., YAMAMOTO, N., KUBOKAWA, I., MORI, T., YANAI, T., HAYAKAWA, A., TAKESHIMA, Y., IJIMA, K., MATSUO, M., NISHIO, H. & NISHIMURA, N. 2012. Rab15 alternative splicing is altered in spheres of neuroblastoma cells. *Oncol Rep*, 27, 2045-9.
- PILS, D., HORAK, P., VANHARA, P., ANEES, M., PETZ, M., ALFANZ, A., GUGERELL, A., WITTINGER, M., GLEISS, A., AUNER, V., TONG, D., ZEILLINGER, R., BRAICU, E. I., SEHOULI, J. & KRAINER, M. 2013. Methylation status of TUSC3 is a prognostic factor in ovarian cancer. *Cancer*, 119, 946-54.
- PROS, E., LANTUEJOUL, S., SANCHEZ-VERDE, L., CASTILLO, S. D., BONASTRE, E., SUAREZ-GAUTHIER, A., CONDE, E., CIGUDOSA, J. C., LOPEZ-RIOS, F., TORRES-LANZAS, J., CASTELLVÍ, J., Y CAJAL, S. R., BRAMBILLA, E. & SANCHEZ-CEPESDES, M. 2013. Determining the profiles and parameters for gene amplification testing of growth factor receptors in lung cancer. *Int J Cancer*.
- PÂQUET, S., FAZLI, L., GROSSE, L., VERREAULT, M., TÊTU, B., RENNIE, P. S., BÉLANGER, A. & BARBIER, O. 2012. Differential expression of the androgen-conjugating UGT2B15 and UGT2B17 enzymes in prostate tumor cells during cancer progression. *J Clin Endocrinol Metab*, 97, E428-32.
- QIU, T., WANG, L., LIU, X. H., WENG, X. D., KUANG, Y. L., CHEN, Z. Y., CHEN, H. & ZHU, H. C. 2012. Over-expressing transporters associated with antigen processing increases antitumor immunity response in prostate cancer. *Cell Immunol*, 279, 167-73.
- ROBERTS, P. J. & STINCHCOMBE, T. E. 2013. KRAS mutation: should we test for it, and does it matter? *J Clin Oncol*, 31, 1112-21.
- ROSELL, R., TARON, M., ALBEROLA, V., MASSUTI, B. & FELIP, E. 2003. Genetic testing for chemotherapy in non-small cell lung cancer. *Lung Cancer*, 41 Suppl 1, S97-102.
- ROZMAN, A., SILAR, M. & KOSNIK, M. 2012. Angiogenin and vascular endothelial growth factor expression in lungs of lung cancer patients. *Radiol Oncol*, 46, 354-9.

- RUSSELL, A. P., WADLEY, G., HESSELINK, M. K., SCHAART, G., LO, S., LÉGER, B., GARNHAM, A., KORNIPS, E., CAMERON-SMITH, D., GIACOBINO, J. P., MUZZIN, P., SNOW, R. & SCHRAUWEN, P. 2003. UCP3 protein expression is lower in type I, IIa and IIx muscle fiber types of endurance-trained compared to untrained subjects. *Pflugers Arch*, 445, 563-9.
- SAKUMA, K., AOKI, M. & KANNAGI, R. 2012. Transcription factors c-Myc and CDX2 mediate E-selectin ligand expression in colon cancer cells undergoing EGF/bFGF-induced epithelial-mesenchymal transition. *Proc Natl Acad Sci U S A*, 109, 7776-81.
- SANCHEZ-PALENCIA, A., GOMEZ-MORALES, M., GOMEZ-CAPILLA, J. A., PEDRAZA, V., BOYERO, L., ROSELL, R. & FÁREZ-VIDAL, M. E. 2011. Gene expression profiling reveals novel biomarkers in nonsmall cell lung cancer. *Int J Cancer*, 129, 355-64.
- SCAGLIOTTI, G., HANNA, N., FOSSELLA, F., SUGARMAN, K., BLATTER, J., PETERSON, P., SIMMS, L. & SHEPHERD, F. A. 2009a. The differential efficacy of pemetrexed according to NSCLC histology: a review of two Phase III studies. *Oncologist*, 14, 253-63.
- SCAGLIOTTI, G., LONGO, M. & NOVELLO, S. 2009b. Nonsmall cell lung cancer in never smokers. *Curr Opin Oncol*, 21, 99-104.
- SCHULZE, D., PLOHMANN, P., HÖBEL, S. & AIGNER, A. 2011. Anti-tumor effects of fibroblast growth factor-binding protein (FGF-BP) knockdown in colon carcinoma. *Mol Cancer*, 10, 144.
- SCHÖNRATH, K., KLEIN-SZANTO, A. J. & BRAUNEWELL, K. H. 2012. The putative tumor suppressor VILIP-1 counteracts epidermal growth factor-induced epidermal-mesenchymal transition in squamous carcinoma cells. *PLoS One*, 7, e33116.
- SEARS, C. R. & TURCHI, J. J. 2012. Complex cisplatin-double strand break (DSB) lesions directly impair cellular non-homologous end-joining (NHEJ) independent of downstream damage response (DDR) pathways. *J Biol Chem*, 287, 24263-72.
- SEGARA, D., BIANKIN, A. V., KENCH, J. G., LANGUSCH, C. C., DAWSON, A. C., SKALICKY, D. A., GOTLEY, D. C., COLEMAN, M. J., SUTHERLAND, R. L. & HENSHALL, S. M. 2005. Expression of HOXB2, a retinoic acid signaling target in pancreatic cancer and pancreatic intraepithelial neoplasia. *Clin Cancer Res*, 11, 3587-96.
- SETHAKORN, N. & DULIN, N. O. 2013. RGS expression in cancer: oncoming the cancer microarray data. *J Recept Signal Transduct Res*, 33, 166-71.
- SHAW, A., YEAP, B., MINO-KENUDSON, M., DIGUMARTHY, S., COSTA, D., HEIST, R., SOLOMON, B., STUBBS, H., ADMANE, S., MCDERMOTT, U., SETTLEMAN, J., KOBAYASHI, S., MARK, E., RODIG, S., CHIRIEAC, L., KWAK, E., LYNCH, T. & IAFRATE, A. 2009. Clinical features and outcome of patients with non-small-cell lung cancer who harbor EML4-ALK. *J Clin Oncol*, 27, 4247-53.
- SHAW, A. T. & ENGELMAN, J. A. 2013. ALK in lung cancer: past, present, and future. *J Clin Oncol*, 31, 1105-11.
- SHAW, R. J. 2009. Tumor suppression by LKB1: SIK-ness prevents metastasis. *Sci Signal*, 2, pe55.
- SHEHATA, M., BIÈCHE, I., BOUTROS, R., WEIDENHOFER, J., FANAYAN, S., SPALDING, L., ZEPS, N., BYTH, K., BRIGHT, R. K., LIDEREAU, R. & BYRNE, J. A. 2008. Nonredundant functions for tumor protein D52-like proteins support specific targeting of TPD52. *Clin Cancer Res*, 14, 5050-60.
- SHOLL, L., YEAP, B., IAFRATE, A., HOLMES-TISCH, A., CHOU, Y., WU, M., GOAN, Y., SU, L., BENEDETTINI, E., YU, J., LODA, M., JÄNNE, P., CHRISTIANI, D. & CHIRIEAC, L. 2009. Lung adenocarcinoma with EGFR amplification has distinct clinicopathologic and molecular features in never-smokers. *Cancer Res*, 69, 8341-8.
- SHUKLA, S., DUAN, S., WU, X., BADNER, J., KASZA, K. & DOLAN, M. 2009. Whole-genome approach implicates CD44 in cellular resistance to carboplatin. *Hum Genomics*, 3, 128-42.
- SIMPSON, C. D., HURREN, R., KASIMER, D., MACLEAN, N., EBERHARD, Y., KETELA, T., MOFFAT, J. & SCHIMMER, A. D. 2012. A genome wide shRNA screen identifies  $\alpha/\beta$

- hydrolase domain containing 4 (ABHD4) as a novel regulator of anoikis resistance. *Apoptosis*, 17, 666-78.
- SMYTH, G. K. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*, 3, Article3.
- SODA, M., CHOI, Y., ENOMOTO, M., TAKADA, S., YAMASHITA, Y., ISHIKAWA, S., FUJIWARA, S., WATANABE, H., KURASHINA, K., HATANAKA, H., BANDO, M., OHNO, S., ISHIKAWA, Y., ABURATANI, H., NIKI, T., SOHARA, Y., SUGIYAMA, Y. & MANO, H. 2007. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature*, 448, 561-6.
- SOH, J., OKUMURA, N., LOCKWOOD, W., YAMAMOTO, H., SHIGEMATSU, H., ZHANG, W., CHARI, R., SHAMES, D., TANG, X., MACAULAY, C., VARELLA-GARCIA, M., VOODER, T., WISTUBA, I., LAM, S., BREKKEN, R., TOYOOKA, S., MINNA, J., LAM, W. & GAZDAR, A. 2009. Oncogene mutations, copy number gains and mutant allele specific imbalance (MASI) frequently occur together in tumor cells. *PLoS One*, 4, e7464.
- STEPHENS, P., HUNTER, C., BIGNELL, G., EDKINS, S., DAVIES, H., TEAGUE, J., STEVENS, C., O'MEARA, S., SMITH, R., PARKER, A., BARTHORPE, A., BLOW, M., BRACKENBURY, L., BUTLER, A., CLARKE, O., COLE, J., DICKS, E., DIKE, A., DROZD, A., EDWARDS, K., FORBES, S., FOSTER, R., GRAY, K., GREENMAN, C., HALLIDAY, K., HILLS, K., KOSMIDOU, V., LUGG, R., MENZIES, A., PERRY, J., PETTY, R., RAINE, K., RATFORD, L., SHEPHERD, R., SMALL, A., STEPHENS, Y., TOFTS, C., VARIAN, J., WEST, S., WIDAA, S., YATES, A., BRASSEUR, F., COOPER, C., FLANAGAN, A., KNOWLES, M., LEUNG, S., LOUIS, D., LOOIJENGA, L., MALKOWICZ, B., PIEROTTI, M., TEH, B., CHENEVIX-TRENCH, G., WEBER, B., YUEN, S., HARRIS, G., GOLDSTRAW, P., NICHOLSON, A., FUTREAL, P., WOOSTER, R. & STRATTON, M. 2004. Lung cancer: intragenic ERBB2 kinase mutations in tumours. *Nature*, 431, 525-6.
- STEVENSON, L., ALLEN, W. L., TURKINGTON, R., JITHESH, P. V., PROUTSKI, I., STEWART, G., LENZ, H. J., VAN SCHAEYBROECK, S., LONGLEY, D. B. & JOHNSTON, P. G. 2012. Identification of galanin and its receptor GalR1 as novel determinants of resistance to chemotherapy and potential biomarkers in colorectal cancer. *Clin Cancer Res*, 18, 5412-26.
- STRATTON, M., CAMPBELL, P. & FUTREAL, P. 2009. The cancer genome. *Nature*, 458, 719-24.
- SU, X., CHAKRAVARTI, D. & FLORES, E. R. 2012. p63 steps into the limelight: crucial roles in the suppression of tumorigenesis and metastasis. *Nat Rev Cancer*, 13, 136-43.
- SUKOWATI, C. H., ROSSO, N., PASCUT, D., ANFUSO, B., TORRE, G., FRANCALANCI, P., CROCÈ, L. S. & TIRIBELLI, C. 2012. Gene and functional up-regulation of the BCRP/ABCG2 transporter in hepatocellular carcinoma. *BMC Gastroenterol*, 12, 160.
- SUN, H. T., CHENG, S. X., TU, Y., LI, X. H. & ZHANG, S. 2013. FoxQ1 Promotes Glioma Cells Proliferation and Migration by Regulating NRXN3 Expression. *PLoS One*, 8, e55693.
- SUN, J. C., LIANG, X. T., PAN, K., WANG, H., ZHAO, J. J., LI, J. J., MA, H. Q., CHEN, Y. B. & XIA, J. C. 2010. High expression level of EDIL3 in HCC predicts poor prognosis of HCC patients. *World J Gastroenterol*, 16, 4611-5.
- SUYAMA, T., SHIRAISHI, T., ZENG, Y., YU, W., PAREKH, N., VESSELLA, R. L., LUO, J., GETZENBERG, R. H. & KULKARNI, P. 2010. Expression of cancer/testis antigens in prostate cancer is associated with disease progression. *Prostate*, 70, 1778-87.
- TAGI, T., MATSUI, T., KIKUCHI, S., HOSHI, S., OCHIAi, T., KOKUBA, Y., KINOSHITA-IDA, Y., KISUMI-HAYASHI, F., MORIMOTO, K., IMAI, T., IMOTO, I., INAZAWA, J. & OTSUJI, E. 2010. Dermokine as a novel biomarker for early-stage colorectal cancer. *J Gastroenterol*, 45, 1201-11.
- TAKANO, S., KATO, Y., YAMAMOTO, T., KANEKO, M. K., ISHIKAWA, E., TSUJIMOTO, Y., MATSUDA, M., NAKAI, K., YANAGIYA, R., MORITA, S., TSUBOI, K. &

- MATSUMURA, A. 2012. Immunohistochemical detection of IDH1 mutation, p53, and internexin as prognostic factors of glial tumors. *J Neurooncol*, 108, 361-73.
- TAKEUCHI, K., SODA, M., TOGASHI, Y., SUZUKI, R., SAKATA, S., HATANO, S., ASAKA, R., HAMANAKA, W., NINOMIYA, H., UEHARA, H., LIM CHOI, Y., SATOH, Y., OKUMURA, S., NAKAGAWA, K., MANO, H. & ISHIKAWA, Y. 2012. RET, ROS1 and ALK fusions in lung cancer. *Nat Med*, 18, 378-81.
- TANIZAKI, J., OKAMOTO, I., SAKAI, K. & NAKAGAWA, K. 2011. Differential roles of trans-phosphorylated EGFR, HER2, HER3, and RET as heterodimerisation partners of MET in lung cancer with MET amplification. *Br J Cancer*, 105, 807-13.
- THUM, T., ERPENBECK, V. J., MOELLER, J., HOHLFELD, J. M., KRUG, N. & BORLAK, J. 2006. Expression of xenobiotic metabolizing enzymes in different lung compartments of smokers and nonsmokers. *Environ Health Perspect*, 114, 1655-61.
- TOMIZAWA, K., SUDA, K., ONOZATO, R., KOSAKA, T., ENDOH, H., SEKIDO, Y., SHIGEMATSU, H., KUWANO, H., YATABE, Y. & MITSUDOMI, T. 2011. Prognostic and predictive implications of HER2/ERBB2/neu gene mutations in lung cancers. *Lung Cancer*, 74, 139-44.
- TOSCHI, L. & CAPPUZZO, F. 2010. Impact of biomarkers on non-small cell lung cancer treatment. *Target Oncol*, 5, 5-17.
- TSAI, J. R., WANG, H. M., LIU, P. L., CHEN, Y. H., YANG, M. C., CHOU, S. H., CHENG, Y. J., YIN, W. H., HWANG, J. J. & CHONG, I. W. 2012. High expression of heme oxygenase-1 is associated with tumor invasiveness and poor clinical outcome in non-small cell lung cancer patients. *Cell Oncol (Dordr)*, 35, 461-71.
- UPADHYAY, R., SANDUJA, S., KAZA, V. & DIXON, D. A. 2013. Genetic polymorphisms in RNA binding proteins contribute to breast cancer survival. *Int J Cancer*, 132, E128-38.
- VADYSIRISACK, D. D., BAENKE, F., ORY, B., LEI, K. & ELLISEN, L. W. 2011. Feedback control of p53 translation by REDD1 and mTORC1 limits the p53-dependent DNA damage response. *Mol Cell Biol*, 31, 4356-65.
- VAINIO, P., MPINDI, J. P., KOHONEN, P., FEY, V., MIRTITI, T., ALANEN, K. A., PERÄLÄ, M., KALLIONIEMI, O. & ILJIN, K. 2012. High-throughput transcriptomic and RNAi analysis identifies AIM1, ERGIC1, TMED3 and TPX2 as potential drug targets in prostate cancer. *PLoS One*, 7, e39801.
- VALLIÈRES, E., SHEPHERD, F. A., CROWLEY, J., VAN HOUTTE, P., POSTMUS, P. E., CARNEY, D., CHANSKY, K., SHAIKH, Z., GOLDSTRAW, P. & INSTITUTIONS, I. A. F. T. S. O. L. C. I. S. C. A. P. 2009. The IASLC Lung Cancer Staging Project: proposals regarding the relevance of TNM in the pathologic staging of small cell lung cancer in the forthcoming (seventh) edition of the TNM classification for lung cancer. *J Thorac Oncol*, 4, 1049-59.
- VAN MEERBEECK, J. P., FENNELL, D. A. & DE RUYSSCHER, D. K. 2011. Small-cell lung cancer. *Lancet*, 378, 1741-55.
- VERDIJK, L. B., KOOPMAN, R., SCHAART, G., MEIJER, K., SAVELBERG, H. H. & VAN LOON, L. J. 2007. Satellite cell content is specifically reduced in type II skeletal muscle fibers in the elderly. *Am J Physiol Endocrinol Metab*, 292, E151-7.
- VIJAYALAKSHMI, R. & KRISHNAMURTHY, A. 2011. Targetable "driver" mutations in non small cell lung cancer. *Indian J Surg Oncol*, 2, 178-88.
- WALKER, M. C., PARRIS, C. N. & MASTERS, J. R. 1987. Differential sensitivities of human testicular and bladder tumor cell lines to chemotherapeutic drugs. *J Natl Cancer Inst*, 79, 213-6.
- WANG, D. & BODOVITZ, S. 2010. Single cell analysis: the new frontier in 'omics'. *Trends Biotechnol*, 28, 281-90.
- WANG, D. & LIPPARD, S. J. 2005. Cellular processing of platinum anticancer drugs. *Nat Rev Drug Discov*, 4, 307-20.

- WANG, L., XIAO, X., LI, D., CHI, Y., WEI, P., WANG, Y., NI, S., TAN, C., ZHOU, X. & DU, X. 2012. Abnormal expression of GADD45B in human colorectal carcinoma. *J Transl Med*, 10, 215.
- WANG, Y., BRODERICK, P., WEBB, E., WU, X., VIJAYAKRISHNAN, J., MATAKIDOU, A., QURESHI, M., DONG, Q., GU, X., CHEN, W., SPITZ, M., EISEN, T., AMOS, C. & HOULSTON, R. 2008. Common 5p15.33 and 6p21.33 variants influence lung cancer risk. *Nat Genet*, 40, 1407-9.
- WATKINS, D., BERMAN, D., BURKHOLDER, S., WANG, B., BEACHY, P. & BAYLIN, S. 2003. Hedgehog signalling within airway epithelial progenitors and in small-cell lung cancer. *Nature*, 422, 313-7.
- WEI, Q., LI, Y. X., LIU, M., LI, X. & TANG, H. 2012. MiR-17-5p targets TP53INP1 and regulates cell proliferation and apoptosis of cervical cancer cells. *IUBMB Life*, 64, 697-704.
- WEIR, B., WOO, M., GETZ, G., PERNER, S., DING, L., BEROUKHIM, R., LIN, W., PROVINCE, M., KRAJA, A., JOHNSON, L., SHAH, K., SATO, M., THOMAS, R., BARLETTA, J., BORECKI, I., BRODERICK, S., CHANG, A., CHIANG, D., CHIRIEAC, L., CHO, J., FUJII, Y., GAZDAR, A., GIORDANO, T., GREULICH, H., HANNA, M., JOHNSON, B., KRIS, M., LASH, A., LIN, L., LINDEMAN, N., MARDIS, E., MCPHERSON, J., MINNA, J., MORGAN, M., NADEL, M., ORRINGER, M., OSBORNE, J., OZENBERGER, B., RAMOS, A., ROBINSON, J., ROTH, J., RUSCH, V., SASAKI, H., SHEPHERD, F., SOUGNEZ, C., SPITZ, M., TSAO, M., TWOMEY, D., VERHAAK, R., WEINSTOCK, G., WHEELER, D., WINCKLER, W., YOSHIZAWA, A., YU, S., ZAKOWSKI, M., ZHANG, Q., BEER, D., WISTUBA, I., WATSON, M., GARRAWAY, L., LADANYI, M., TRAVIS, W., PAO, W., RUBIN, M., GABRIEL, S., GIBBS, R., VARMUS, H., WILSON, R., LANDER, E. & MEYERSON, M. 2007. Characterizing the cancer genome in lung adenocarcinoma. *Nature*, 450, 893-8.
- WENG, H. Y., HUANG, H. L., ZHAO, P. P., ZHOU, H. & QU, L. H. 2012. Translational repression of cyclin D3 by a stable G-quadruplex in its 5' UTR: implications for cell cycle regulation. *RNA Biol*, 9, 1099-109.
- WHIPPLE, C. A., LANDER, A. D. & KORC, M. 2008. Discovery of a novel molecule that regulates tumor growth and metastasis. *ScientificWorldJournal*, 8, 1250-3.
- WISTUBA, I. I., BRYANT, D., BEHRENS, C., MILCHGRUB, S., VIRMANI, A. K., ASHFAQ, R., MINNA, J. D. & GAZDAR, A. F. 1999. Comparison of features of human lung cancer cell lines and their corresponding tumors. *Clin Cancer Res*, 5, 991-1000.
- WOJTALLA, A. & ARCARO, A. 2011. Targeting phosphoinositide 3-kinase signalling in lung cancer. *Crit Rev Oncol Hematol*, 80, 278-90.
- WU, Z., WANG, H., CHU, X., CHEN, J. & FANG, S. 2013. Association between FAS-1377 G/A polymorphism and susceptibility to gastric cancer: evidence from a meta-analysis. *Tumour Biol*, 34, 2147-52.
- WUTTIG, D., ZASTROW, S., FÜSSEL, S., TOMA, M. I., MEINHARDT, M., KALMAN, K., JUNKER, K., SANJMYATAV, J., BOLL, K., HACKERMÜLLER, J., ROLLE, A., GRIMM, M. O. & WIRTH, M. P. 2012. CD31, EDNRB and TSPAN7 are promising prognostic markers in clear-cell renal cell carcinoma revealed by genome-wide expression analyses of primary tumors and metastases. *Int J Cancer*, 131, E693-704.
- XIE, C., GOU, M. L., YI, T., DENG, H., LI, Z. Y., LIU, P., QI, X. R., HE, X., WEI, Y. & ZHAO, X. 2011. Efficient inhibition of ovarian cancer by truncation mutant of FILIP1L gene delivered by novel biodegradable cationic heparin-polyethyleneimine nanogels. *Hum Gene Ther*, 22, 1413-22.
- YAMAMOTO, H., SHIGEMATSU, H., NOMURA, M., LOCKWOOD, W., SATO, M., OKUMURA, N., SOH, J., SUZUKI, M., WISTUBA, I., FONG, K., LEE, H., TOYOOKA, S., DATE, H., LAM, W., MINNA, J. & GAZDAR, A. 2008. PIK3CA mutations and copy number gains in human lung cancers. *Cancer Res*, 68, 6913-21.

- YANG, D., KEDEI, N., LI, L., TAO, J., VELASQUEZ, J. F., MICHALOWSKI, A. M., TÓTH, B. I., MARINCÁS, R., VARGA, A., BÍRÓ, T., YUSPA, S. H. & BLUMBERG, P. M. 2010. RasGRP3 contributes to formation and maintenance of the prostate cancer phenotype. *Cancer Res*, 70, 7905-17.
- YANG, D., TAO, J., LI, L., KEDEI, N., TÓTH, Z. E., CZAP, A., VELASQUEZ, J. F., MIHOVA, D., MICHALOWSKI, A. M., YUSPA, S. H. & BLUMBERG, P. M. 2011. RasGRP3, a Ras activator, contributes to signaling and the tumorigenic phenotype in human melanoma. *Oncogene*, 30, 4590-600.
- YANG, H. J., KIM, N., SEONG, K. M., YOUN, H. & YOUN, B. 2013. Investigation of radiation-induced transcriptome profile of radioresistant non-small cell lung cancer A549 cells using RNA-seq. *PLoS One*, 8, e59319.
- YE, J., WU, D., SHEN, J., WU, P., NI, C., CHEN, J., ZHAO, J., ZHANG, T., WANG, X. & HUANG, J. 2012. Enrichment of colorectal cancer stem cells through epithelial-mesenchymal transition via CDH1 knockdown. *Mol Med Rep*, 6, 507-12.
- YILDIRIM, S., ALTUN, S., GUMUSHAN, H., PATEL, A. & DJAMGOZ, M. B. 2012. Voltage-gated sodium channel activity promotes prostate cancer metastasis in vivo. *Cancer Lett*, 323, 58-61.
- YOUNG, C. D., PFEFFERLE, A. D., OWENS, P., KUBA, M. G., REXER, B. N., BALKO, J. M., SÁNCHEZ, V., CHENG, H., PEROU, C. M., ZHAO, J. J., COOK, R. S. & ARTEAGA, C. L. 2013. Conditional loss of ErbB3 delays mammary gland hyperplasia induced by mutant PIK3CA without affecting mammary tumor latency, gene expression, or signaling. *Cancer Res*, 73, 4075-85.
- YUAN, H., KAJIYAMA, H., ITO, S., YOSHIKAWA, N., HYODO, T., ASANO, E., HASEGAWA, H., MAEDA, M., SHIBATA, K., HAMAGUCHI, M., KIKKAWA, F. & SENG, T. 2013. ALX1 Induces Snail Expression to Promote Epithelial-to-Mesenchymal Transition and Invasion of Ovarian Cancer Cells. *Cancer Res*.
- ZELLEN, M. 1973. Keynote address on biostatistics and data retrieval. *Cancer Chemother Rep* 3, 4, 31-42.
- ZHANG, H., XU, X., GAJEWIAK, J., TSUKAHARA, R., FUJIWARA, Y., LIU, J., FELS, J. I., PERYGIN, D., PARRILL, A. L., TIGYI, G. & PRESTWICH, G. D. 2009. Dual activity lysophosphatidic acid receptor pan-antagonist/autotaxin inhibitor reduces breast cancer cell migration in vitro and causes tumor regression in vivo. *Cancer Res*, 69, 5441-9.
- ZHANG, J., PARK, S. I., ARTIME, M. C., SUMMY, J. M., SHAH, A. N., BOMSER, J. A., DORFLEUTNER, A., FLYNN, D. C. & GALLICK, G. E. 2007. AFAP-110 is overexpressed in prostate cancer and contributes to tumorigenic growth by regulating focal contacts. *J Clin Invest*, 117, 2962-73.
- ZHANG, X., WU, Q., GAN, L., YU, G. Z., WANG, R., WANG, Z. S., WANG, J. J. & WANG, X. 2013a. Reduced group IVA phospholipase A2 expression is associated with unfavorable outcome for patients with gastric cancer. *Med Oncol*, 30, 454.
- ZHANG, X., ZHANG, S., YANG, X., YANG, J., ZHOU, Q., YIN, L., AN, S., LIN, J., CHEN, S., XIE, Z., ZHU, M. & WU, Y. L. 2010. Fusion of EML4 and ALK is associated with development of lung adenocarcinomas lacking EGFR and KRAS mutations and is correlated with ALK expression. *Mol Cancer*, 9, 188.
- ZHANG, Y., TAKAHASHI, S., TASAKA, A., YOSHIMA, T., OCHI, H. & CHAYAMA, K. 2013b. Involvement of microRNA-224 in cell proliferation, migration, invasion, and anti-apoptosis in hepatocellular carcinoma. *J Gastroenterol Hepatol*, 28, 565-75.
- ZHENG, F., HE, K., LI, X., ZHAO, D., SUN, F., ZHANG, Y., NIE, D., CHU, W., SUN, Y. & LU, Y. 2013. Transient over-expression of TGFBR3 induces apoptosis in human nasopharyngeal carcinoma CNE-2Z cells. *Biosci Rep*.
- ZIMMERMANN, M., ARACHCHIGE-DON, A. S., DONALDSON, M. S., DALLAPIAZZA, R. F., COWAN, C. E. & HORNE, M. C. 2012. Elevated cyclin G2 expression intersects with DNA

damage checkpoint signaling and is required for a potent G2/M checkpoint arrest response to doxorubicin. *J Biol Chem*, 287, 22838-53.

ZONG, C., LU, S., CHAPMAN, A. R. & XIE, X. S. 2012. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science*, 338, 1622-6.

## **9 Appendices**

### **9.1 E-Appendices for use in conjunction with Chapter 4**

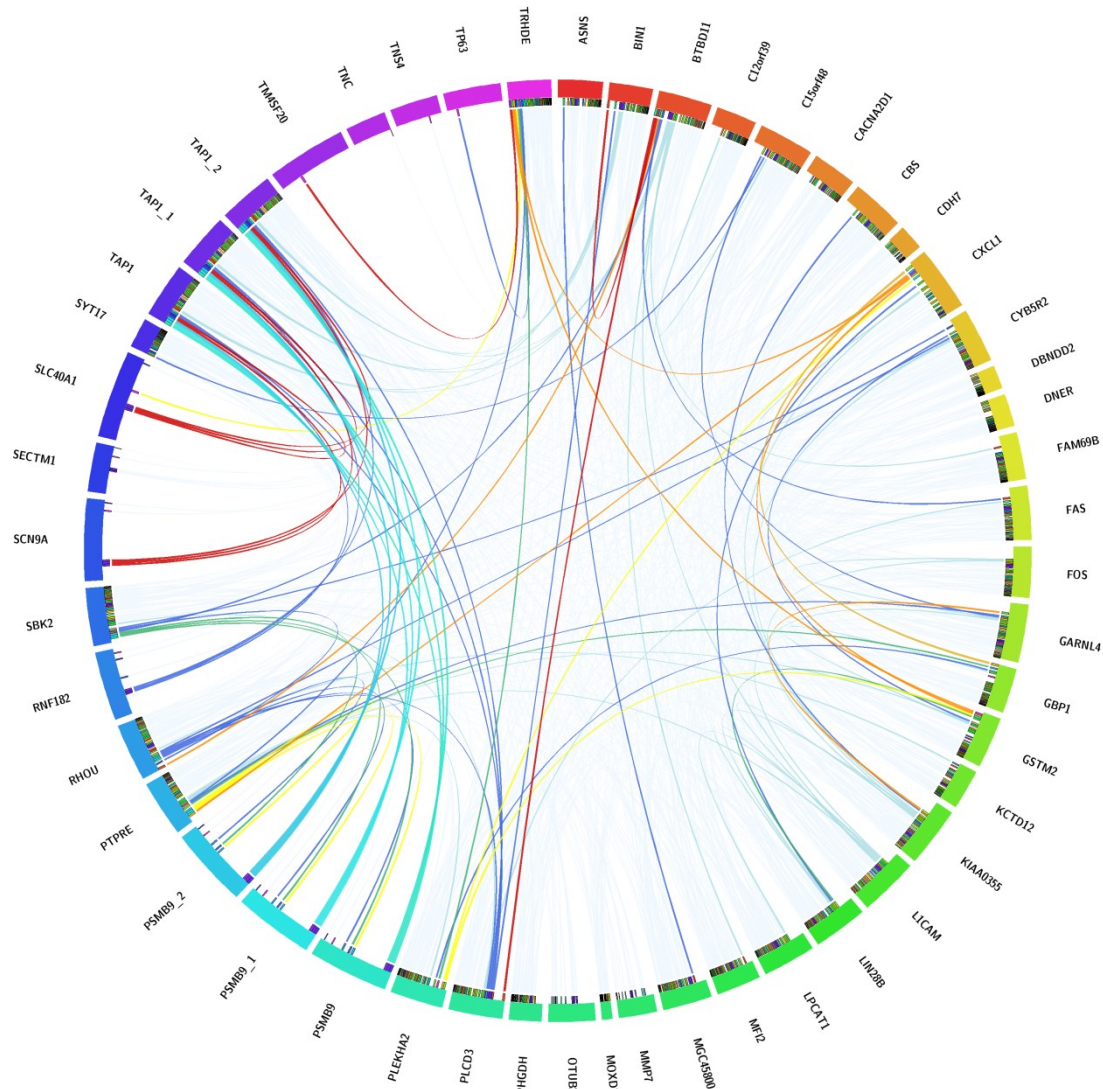
- 9.1.1 Differentially Expressed Genes from the IC50 Dose 3 Week Recovery Time course B-H adjusted p-value  $<1E-14$
- 9.1.2 Differentially Expressed Genes from the IC25 Dose 1 Week Recovery Time course B-H adjusted p-value  $<1E-14$
- 9.1.3 Genes from the Temporal Expression Clusters of Interest, IC50 Dose 3 Week Recovery Time course
- 9.1.4 Genes from the Temporal Expression Clusters of Interest, IC25 Dose 1 Week Recovery Time course
- 9.1.5 375 Genes of Overlap found in to be Differentially Expressed in Both Treatment Time Courses but not in the Normal Untreated Time Course

### **9.2 E-Appendices for use in conjunction with Chapter 5**

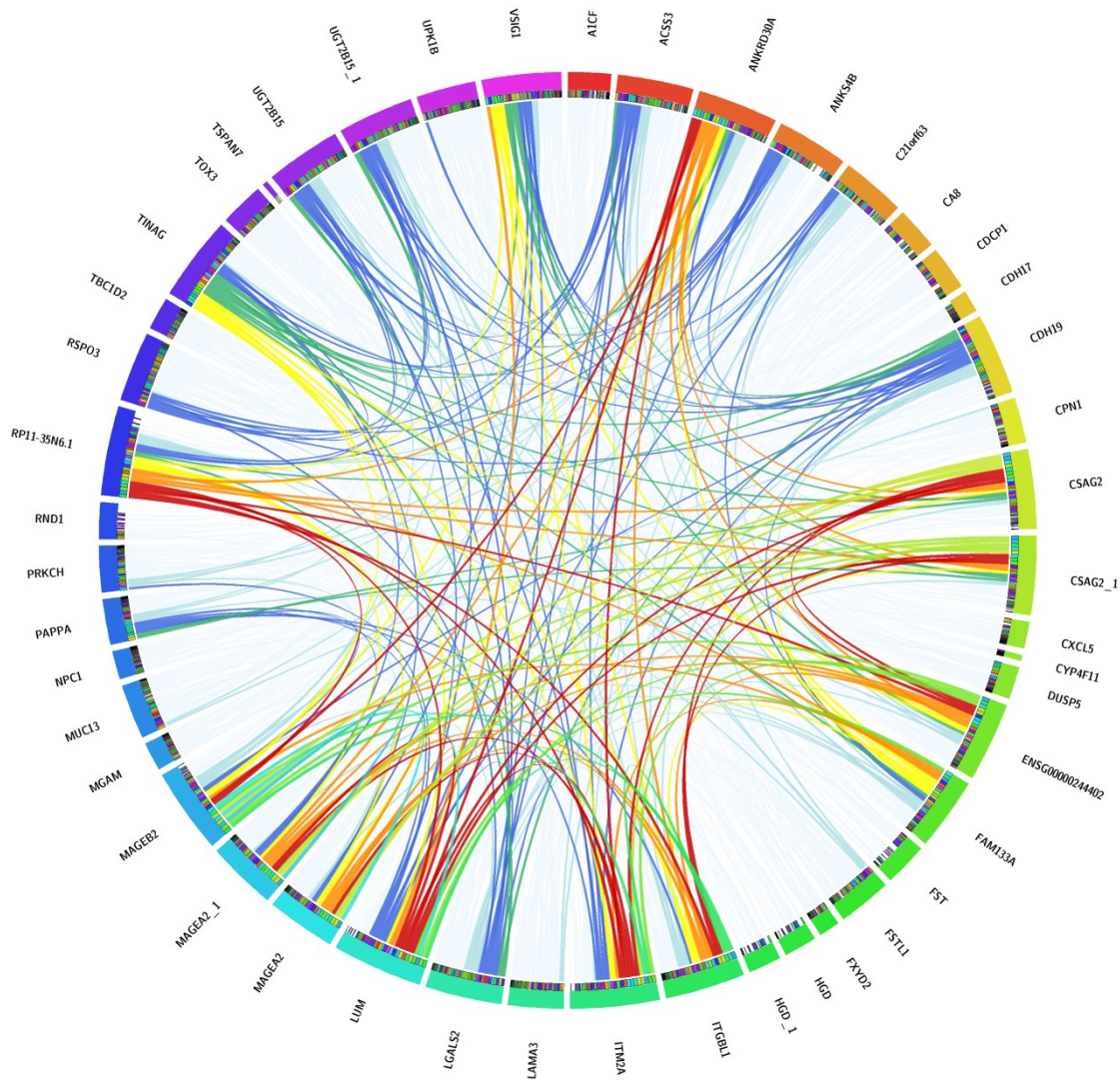
- 9.2.1 Gene Membership for WGCNA Modules from the IC50 Dose 3 Week Recovery Time Course
- 9.2.2 Genes Overlapping WGCNA Modules from the IC50 Dose 3 Week Recovery Time Course and Temporal Expression Clusters of Interest
- 9.2.3 Gene Membership for WGCNA Modules from the IC25 Dose 1 Week Recovery Time Course
- 9.2.4 Genes Overlapping WGCNA Modules from the IC25 Dose 1 Week Recovery Time Course and Temporal Expression Clusters of Interest
- 9.2.5 Gene Membership for the Reduced WGCNA Modules from the IC25 Dose 1 Week Recovery Time Course
- 9.2.6 Genes Overlapping Reduced WGCNA Modules from the IC25 Dose 1 Week Recovery Time Course and Temporal Expression Clusters of Interest
- 9.2.7 Gene Membership for WGCNA Modules from the Normal Untreated Time Course
- 9.2.8 Gene Membership for WGCNA Modules from 375 Genes Differentially Expressed in both Treatment Time Courses but not in the Normal Untreated Time Course

## 9.3 Circos Plots of Gene Interconnectivity within Modules

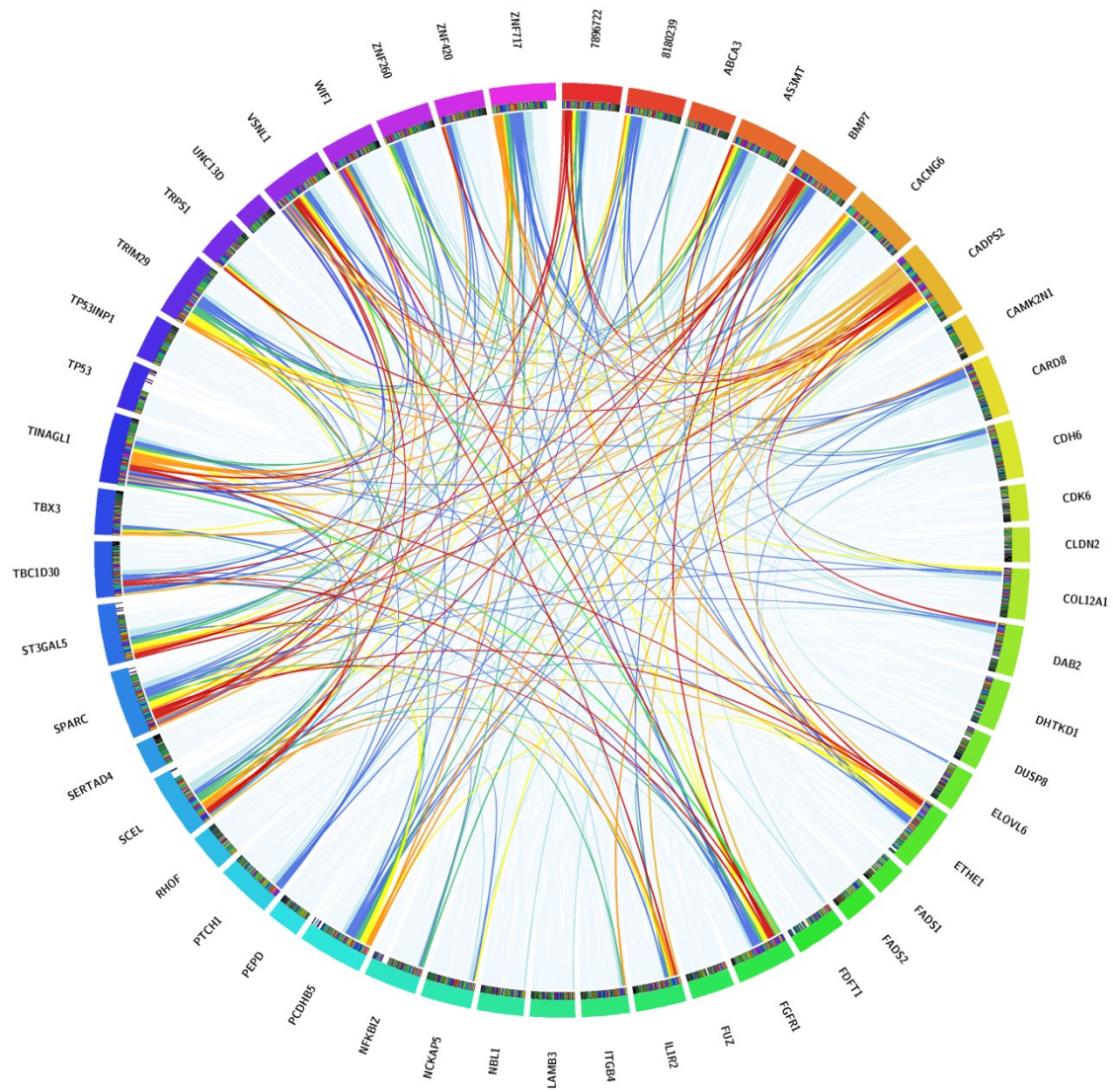
### 9.3.1 IC50 Dose 3 Week Recovery Time Course



**Figure 1** Circos plot of the all 50 genes in the Pink module from the IC50 dose 3 week recovery time course. Some strong connections are visible between the genes although the majority of the connections appear to be pale blue and consequently are less well correlated connections. Genes that appear to be highly connected include *BIN1*, *BTBD11*, *CXCL1*, *GARNL4*, *GSTM2*, *PLCD3*, *PTPRE*, *RHOA*, *SCN9A*, *SLC40A1*, *TAP1* (three transcripts for this gene in the module) and *TRHDE*.

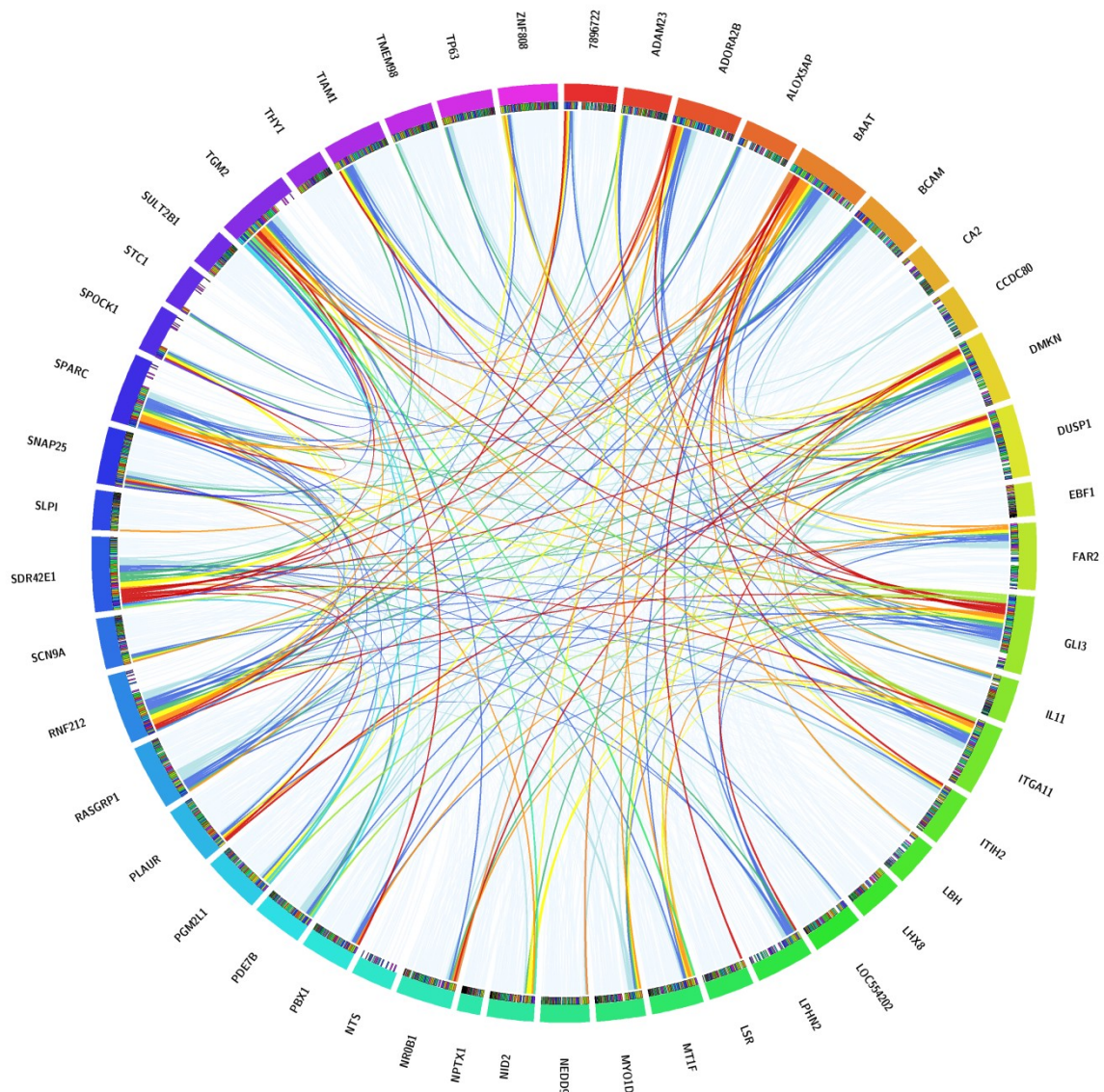


**Figure 2** Circos plot of the all 46 genes in the Green module from the IC50 dose 3 week recovery. Genes that appear to be highly connected include *ANKRD30A*, *CSAG2* (two transcripts), *ENSG00000244402*, *ITGBL1*, *ITM2A*, *LUM*, *MAGEA2* (two transcripts), *MAGEB2*, *RP11-35N6.1*, and *VSIG1*.

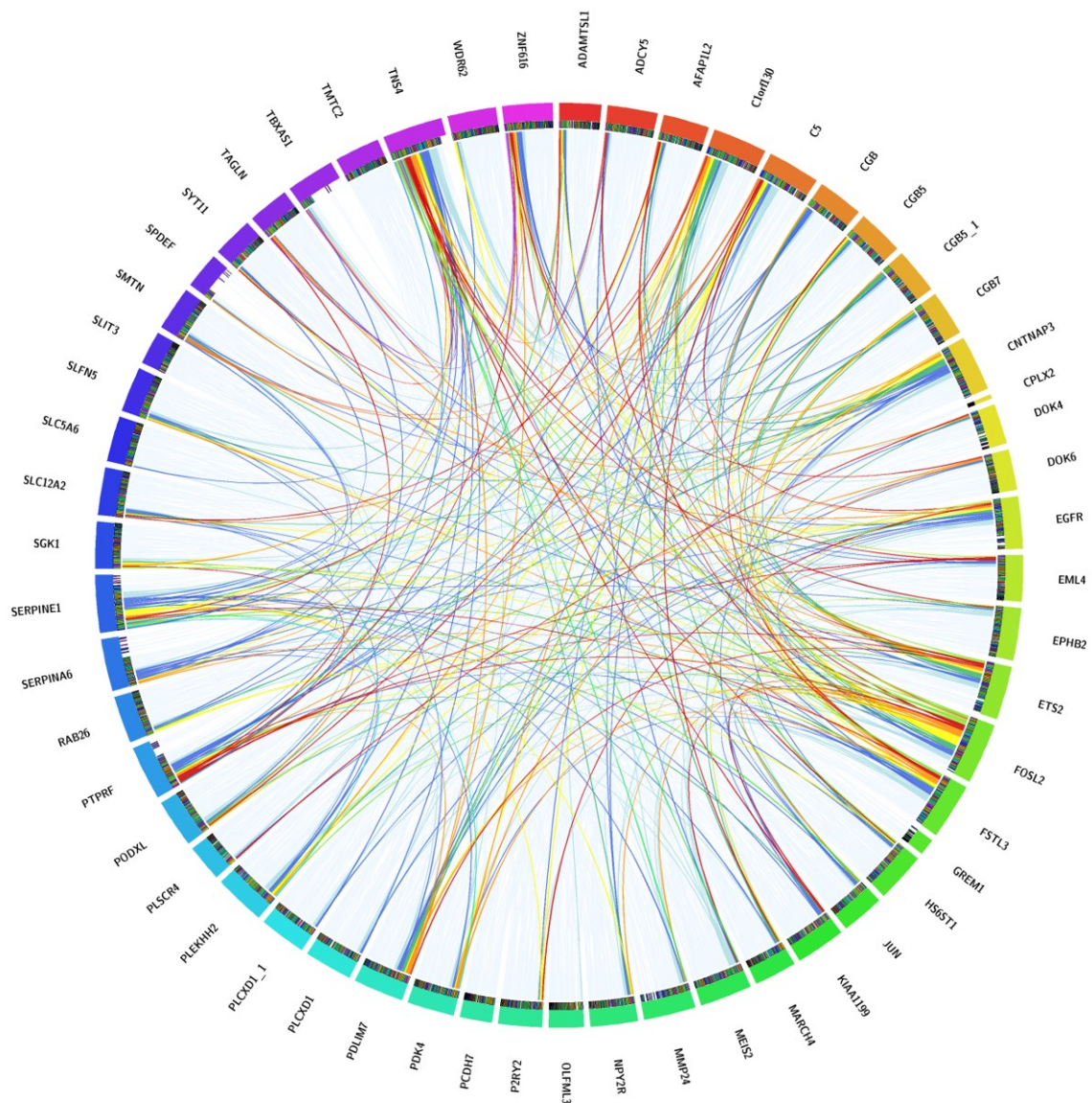


**Figure 3** Circos plot of the top 50 genes in the Yellow module from the IC50 dose 3 week recovery. Genes that appear to be highly connected include *BMP7*, *CADNG6*, *ETHE1*, *FGFR1*, *SCEL*, *SPARC*, *ST3GAL5*, *TINAGL1*, *TRIM29*, *VSNL1*, and *ZNF717*.

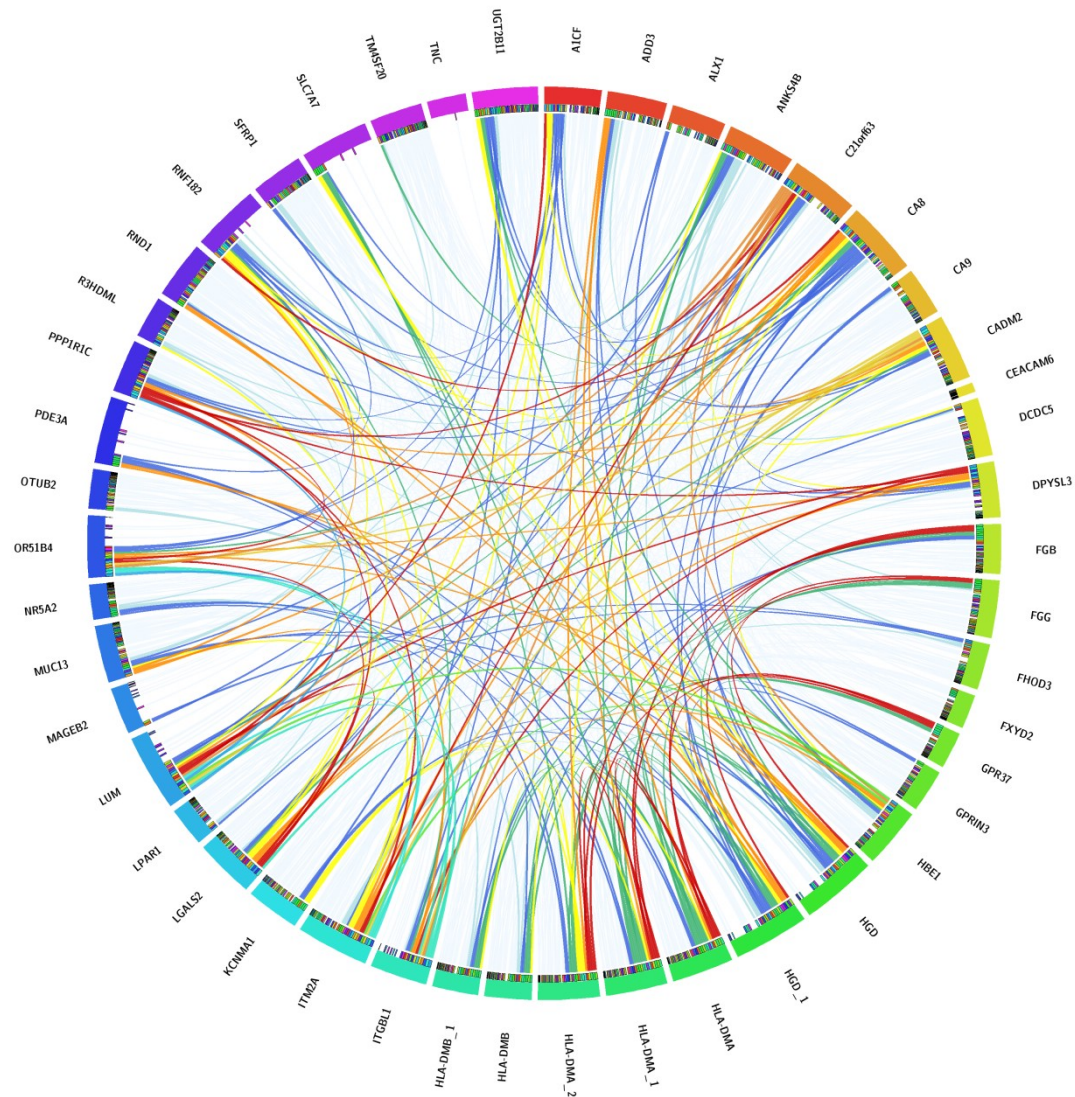
### 9.3.2 IC25 Dose 1 Week Recovery Time Course



**Figure 4** Circos plot of the Brown module from the IC25 Dose 1 Week Recovery Time Course. Strong connections are visible between the 48 genes on the plot. Highly connected genes include *BAAT*, *DMKN*, *DUSP1*, *GLI3*, *ITGA11*, *RNF212*, *SDR42E1*, *SPARC*, and *TGM2*. All these transcripts have multiple connections to other transcripts, some of which are high in the connectivity levels.

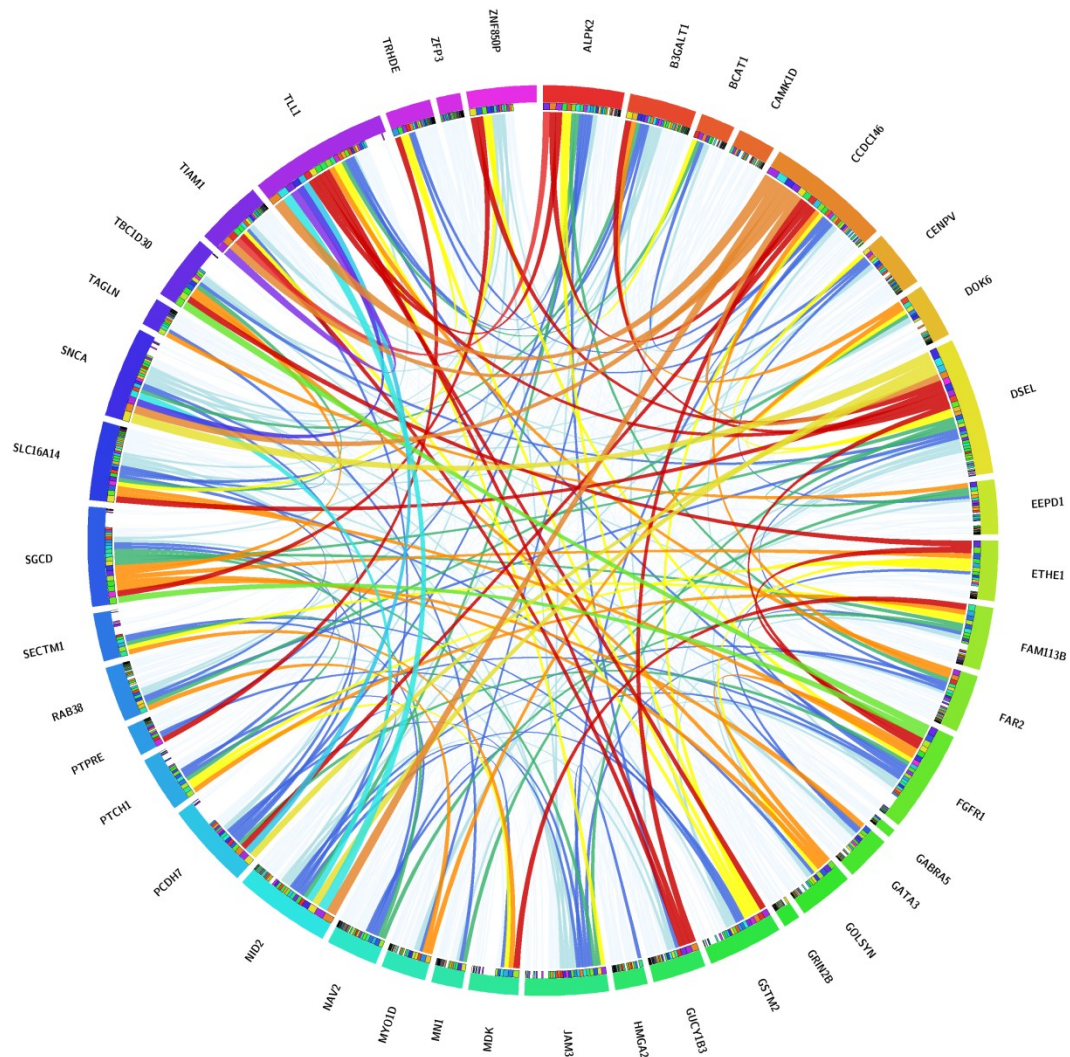


**Figure 5** Circos plot of the Blue module from the IC25 Dose 1 Week Recovery Time Course. Strong connections are visible between the 55 genes on the plot. Highly connected genes include *C1orf130*, *C5*, *EGFR*, *EML4*, *ETS2*, *FOSL2*, *FSTL3*, *PDLIM7*, *PTPRF*, and *TNS4*. All these transcripts have multiple connections to other transcripts, some of which are high in connectivity level.

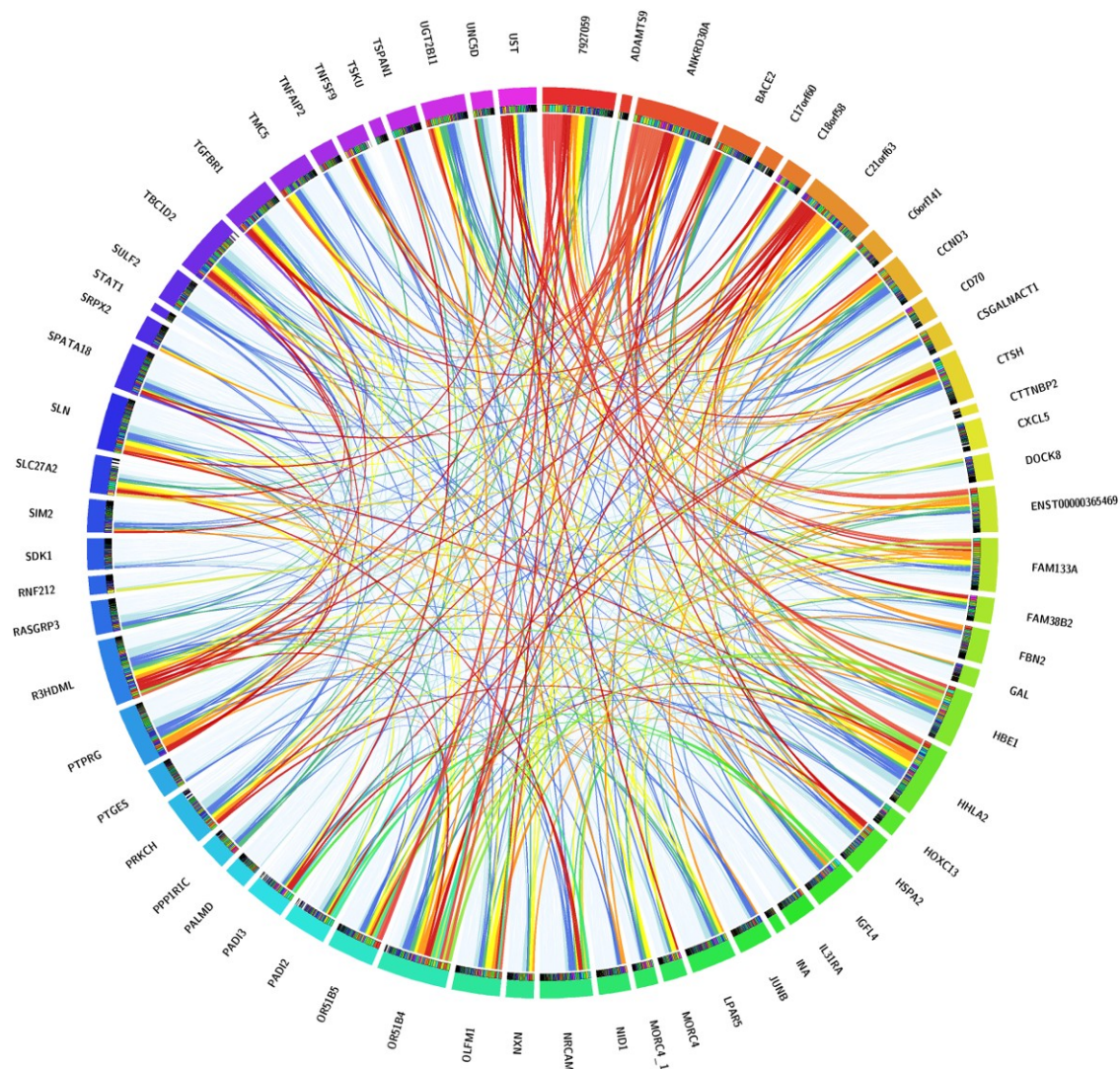


**Figure 6** Circos plot of the yellow module from the IC25 Dose 1 Week Recovery Time Course. Strong connections are visible between the 46 genes on the plot. Highly connected genes include *C21orf163*, *CA8*, *DPYSL3*, *HLA-DMA* (three transcripts), *ITM2A*, *LGALS2*, *LUM*, *OR51B4*, *PPP1R1C*, and *RNF182*. All these transcripts have multiple connections to other transcripts, some of which are high in connectivity level.

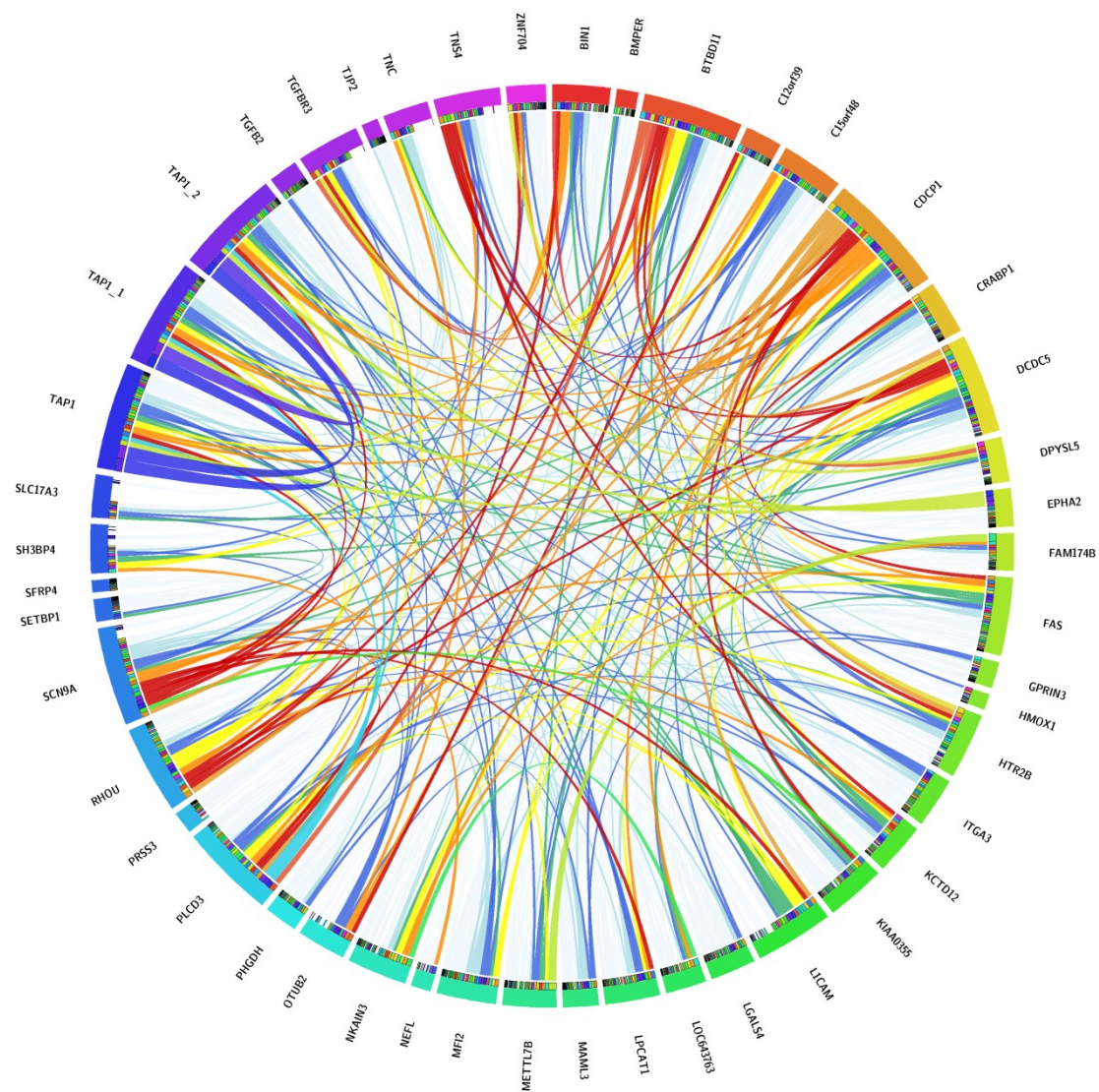
### 9.3.3 375 Overlapping Genes between the two treated time courses



**Figure 7** Circos plot of the Green module from the 375 genes only differentially expressed in the treated time courses using the IC50 dose 3 week recovery time course data. Strong connections are visible across all 41 genes. Despite this, certain hubs with several, strongly correlated connections appear prominent, such as *CCDC146*, *DSEL*, *FGFR1*, and *TLL1*.



**Figure 8** Circos plot of the Blue module from the 375 genes only differentially expressed in the treated time courses using the IC50 dose 3 week recovery time course data. Strong connections are visible across all 65 genes.



**Figure 9** Circos plot of the yellow module from the 375 genes only differentially expressed in the treated time courses using the IC50 dose 3 week recovery time course data. Strong connections are visible across all 46 genes. Despite this, certain hubs with several, strongly correlated connections appear prominent, such as *BTBD11*, *CDCP1*, *DCDC5*, *FAS*, *LICAM*, *RHOU*, *SCN9A*, and *TNS4*.

## 9.4 Copyright Permissions

| Page Number | Type of work: text, figure, map, etc. | Source work                                              | Copyright holder & year                                            | Work out of copyright | Permission to re-use | Permission requested | Permission refused | Orphan work |
|-------------|---------------------------------------|----------------------------------------------------------|--------------------------------------------------------------------|-----------------------|----------------------|----------------------|--------------------|-------------|
| Page 23     | Table                                 | Journal of Thoracic Oncology (2007), Aug; 2(8): 706-14   | © Goldstraw et al., 2007 published by Journal of Thoracic Oncology |                       | ✓                    |                      |                    |             |
| Page 27     | Figure                                | New England Journal of Medicine (2008), vol 359, 1367-80 | © Herbst et al., 2008 published by Massachusetts Medical Society   |                       | ✓                    |                      |                    |             |
| Page 30     | Figure                                | ErbB/HER signal pathway<br>cellsignal.com                | © Cell Signaling Technologies Inc.                                 |                       | ✓                    |                      |                    |             |
| Page 31     | Figure                                | New England Journal of Medicine (2008), vol 359, 1367-80 | © Herbst et al., 2008 published by Massachusetts Medical Society   |                       | ✓                    |                      |                    |             |
| Page 40     | Figure                                | Nature Reviews Cancer (2001), vol 1, 22-33               | © Errol C. Friedberg, 2001 published by Macmillan Publishers Ltd   |                       | ✓                    |                      |                    |             |
| Page 47     | Figure                                | PLoS One (2009), Vol 4, e7464                            | Open Access<br>Soh et al., 2009                                    |                       | ✓                    |                      |                    |             |

|                |               |                                                                    |                                                                                       |  |   |  |  |  |
|----------------|---------------|--------------------------------------------------------------------|---------------------------------------------------------------------------------------|--|---|--|--|--|
| <b>Page 52</b> | <b>Figure</b> | <b>Nature Reviews Drug<br/>Discovery (2005), vol<br/>4, 307-20</b> | <b>© Wang and<br/>Lippard, 2005<br/>published by<br/>Macmillan<br/>Publishers Ltd</b> |  | ✓ |  |  |  |
|----------------|---------------|--------------------------------------------------------------------|---------------------------------------------------------------------------------------|--|---|--|--|--|

**WOLTERS KLUWER HEALTH LICENSE  
TERMS AND CONDITIONS**

Mar 20, 2014

This is a License Agreement between Elizabeth J Richards ("You") and Wolters Kluwer Health ("Wolters Kluwer Health") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Wolters Kluwer Health, and the payment terms and conditions.

**All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.**

|                                        |                                                                                                                                                                                  |
|----------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| License Number                         | 3343650866866                                                                                                                                                                    |
| License date                           | Mar 07, 2014                                                                                                                                                                     |
| Licensed content publisher             | Wolters Kluwer Health                                                                                                                                                            |
| Licensed content publication           | Journal of Thoracic Oncology                                                                                                                                                     |
| Licensed content title                 | The IASLC Lung Cancer Staging Project: Proposals for the Revision of the TNM Stage Groupings in the Forthcoming (Seventh) Edition of the TNM Classification of Malignant Tumours |
| Licensed content author                | Peter Goldstraw, John Crowley, Kari Chansky, et al                                                                                                                               |
| Licensed content date                  | Jan 1, 2007                                                                                                                                                                      |
| Volume Number                          | 2                                                                                                                                                                                |
| Issue Number                           | 8                                                                                                                                                                                |
| Type of Use                            | Dissertation/Thesis                                                                                                                                                              |
| Requestor type                         | Individual                                                                                                                                                                       |
| Portion                                | Figures/table/illustration                                                                                                                                                       |
| Number of figures/tables/illustrations | 1                                                                                                                                                                                |
| Figures/tables/illustrations used      | Table 8. Proposed TNM Stage Groupings                                                                                                                                            |
| Author of this Wolters Kluwer article  | No                                                                                                                                                                               |
| Title of your thesis / dissertation    | Molecular Profiling of Lung Cancer                                                                                                                                               |
| Expected completion date               | Mar 2014                                                                                                                                                                         |
| Estimated size(pages)                  | 240                                                                                                                                                                              |
| Billing Type                           | Invoice                                                                                                                                                                          |
| Billing address                        | 22B Putney High Street<br>London, SW15 1SL<br>United Kingdom                                                                                                                     |
| Total                                  | 0.00 GBP                                                                                                                                                                         |
| Terms and Conditions                   |                                                                                                                                                                                  |

**Terms and Conditions**

1. A credit line will be prominently placed and include: for books - the author(s), title of book, editor, copyright holder, year of publication; For journals - the author(s), title of article,

- title of journal, volume number, issue number and inclusive pages.
2. The requestor warrants that the material shall not be used in any manner which may be considered derogatory to the title, content, or authors of the material, or to Wolters Kluwer.
  3. Permission is granted for a one time use only within 12 months from the date of this invoice. Rights herein do not apply to future reproductions, editions, revisions, or other derivative works. Once the 12-month term has expired, permission to renew must be submitted in writing.
  4. Permission granted is non-exclusive, and is valid throughout the world in the English language and the languages specified in your original request.
  5. Wolters Kluwer cannot supply the requestor with the original artwork or a "clean copy."
  6. The requestor agrees to secure written permission from the author (for book material only).
  7. Permission is valid if the borrowed material is original to a Wolters Kluwer imprint (Lippincott-Raven Publishers, Williams & Wilkins, Lea & Febiger, Harwal, Igaku-Shoin, Rapid Science, Little Brown & Company, Harper & Row Medical, American Journal of Nursing Co, and Urban & Schwarzenberg - English Language).
  8. If you opt not to use the material requested above, please notify Rightslink within 90 days of the original invoice date.
  9. Please note that articles in the ahead-of-print stage of publication can be cited and the content may be re-used by including the date of access and the unique DOI number. Any final changes in manuscripts will be made at the time of print publication and will be reflected in the final electronic version of the issue. **Disclaimer:** Articles appearing in the Published Ahead-of-Print section have been peer-reviewed and accepted for publication in the relevant journal and posted online before print publication. Articles appearing as publish ahead-of-print may contain statements, opinions, and information that have errors in facts, figures, or interpretation. Accordingly, Lippincott Williams & Wilkins, the editors and authors and their respective employees are not responsible or liable for the use of any such inaccurate or misleading data, opinion or information contained in the articles in this section.
  10. This permission does not apply to images that are credited to publications other than Wolters Kluwer journals. For images credited to non-Wolters Kluwer journal publications, you will need to obtain permission from the journal referenced in the figure or table legend or credit line before making any use of the image(s) or table(s).
  11. In case of **Disease Colon Rectum, Plastic Reconstructive Surgery, The Green Journal, Critical Care Medicine, Pediatric Critical Care Medicine, the American Heart Publications, the American Academy of Neurology** the following guideline applies: no drug brand/trade name or logo can be included in the same page as the material re-used
  12. When requesting a permission to translate a full text article, Wolters Kluwer/Lippincott Williams & Wilkins requests to receive the pdf of the translated document
  13. "Adaptations of single figures do not require Wolters Kluwer **further** approval if the permission has been granted previously. However, the adaptation should be credited as follows: ?Adapted with permission from Lippincott Williams and Wilkins/Wolters Kluwer Health: [JOURNAL NAME] (reference citation), copyright (year of publication)" **Please note that modification of text within figures or full-text articles is strictly forbidden.**
  14. The following statement needs to be added when reprinting the material in Open Access journals only: 'promotional and commercial use of the material in print, digital or mobile device format is prohibited without the permission from the publisher Lippincott Williams & Wilkins. Please contact journalpermissions@lww.com for further information".
  15. Other Terms and Conditions:

v1.8

**If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK501244820.**

**Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.**

**Make Payment To:  
Copyright Clearance Center  
Dept 001**

**P.O. Box 843006  
Boston, MA 02284-3006**

**For suggestions or comments regarding this order, contact RightsLink Customer Support: [customercare@copyright.com](mailto:customercare@copyright.com) or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.**

**Gratis licenses (referencing \$0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.**

---

---

**RE: Permission to reproduce two images from an NEJM paper in a PhD thesis?**

Permissions [permissions@nejm.org]

**Sent:** 07 March 2014 14:36

**To:** Richards, Elizabeth

Dear Ms. Richards,

Thank you for your email. Yes, figures and tables may be used in a thesis or dissertation.

Sincerely,

*Jennifer Moran*

Jennifer Moran | *Licensing & Permissions Specialist* | NEJM Group  
860 Winter Street, Waltham, MA 02451 | 781-434-7382 | [jmoran@mms.org](mailto:jmoran@mms.org)



---

**From:** Richards, Elizabeth [mailto:elizabeth.richards08@imperial.ac.uk]

**Sent:** Friday, March 07, 2014 6:43 AM

**To:** Permissions

**Subject:** Permission to reproduce two images from an NEJM paper in a PhD thesis?

**Importance:** High

Hi,

I would like to check that the following statement from your permissions website covers figures as well as text content.

Content (full-text or portions thereof) may be used in print and electronic versions of your dissertation or thesis without formal permission from the Massachusetts Medical Society, Publisher of the New England Journal of Medicine.

**The following credit line must be printed along with the copyrighted material:**

"Reproduced with permission from (scientific reference citation), Copyright Massachusetts Medical Society.

If you could confirm this as soon as possible, I would be most grateful.

Yours sincerely,  
Elizabeth Richards

This email message is a private communication. The information transmitted, including attachments, is intended only for the person or entity to which it is addressed and may contain confidential, privileged, and/or proprietary material. Any review, duplication, retransmission, distribution, or other use of, or taking of any action in reliance upon, this information by persons or entities other than the intended recipient is unauthorized by the sender and is prohibited. If you have received this message in error, please contact the sender immediately by return email and delete the original message from all computer systems. Thank you.

**Re: [Request ID :##437403##] : Request for reproduction of an image from CellSignal.com in a PhD thesis**

marketing@cellsignal.com on behalf of Marketing [marketing@cellsignal.net]

**Sent:** 13 March 2014 00:24

**To:** Richards, Elizabeth

**Cc:** info@cellsignal.com; support@cellsignal.com; webmaster@cellsignal.com

Hi Elizabeth,

Thank you for your interest in Cell Signaling Technology. We welcome the use of our graphics for academic and educational use. As long as it is not being used for commercial use (you will need special permission) we would only require for you to acknowledge us with the text "Illustration reproduced courtesy of Cell Signaling Technology, Inc. ([www.cellsignal.com](http://www.cellsignal.com))." that is clearly represented with the image in all media. Please let me know if you require anything else.

Kind regards,

Aileen

---

Aileen Taylor

Marketing

Cell Signaling Technology

3 Trask Lane

Danvers, MA 01923

office 978-867-2420

ataylor@cellsignal.com

On Mar 7, 2014 06:30 AM Richards, Elizabeth wrote ---

Dear Cell Signaling Technology, Inc.

I have completed my PhD thesis at Imperial College London entitled 'Molecular Profiling of Lung Cancer'.

I seek your permission to reprint, in my thesis an image from:

<http://www.cellsignal.com/reference/pathway/>.

The image to be reproduced is: ErbB/HER Signaling

([http://www.cellsignal.com/reference/pathway/ErbB\\_HER.html](http://www.cellsignal.com/reference/pathway/ErbB_HER.html))

I would like to include the image in the printed examination copy of my thesis and also the electronic version which will be added to Spiral, Imperial's online repository

<http://spiral.imperial.ac.uk/> and made available to the public under a [Creative Commons Attribution-NonCommercial-NoDerivs licence](#).

If you are happy to grant me all the permissions requested, please return this email. If you wish to grant only some of the permissions requested, please list these below.

Yours sincerely,

Elizabeth Richards

**Permission granted for the use requested above:**

I confirm that I am the copyright holder of the extract above and hereby give permission to include it in the print and electronic version of your thesis. I understand that the electronic version of the thesis will be made available, via the internet, for non-commercial purposes

under the terms of the user licence.

**[please edit the text above if you wish to grant more specific permission]**

Signed:

Name:

Organisation:

Job title:

**NATURE PUBLISHING GROUP LICENSE  
TERMS AND CONDITIONS**

Mar 20, 2014

This is a License Agreement between Elizabeth J Richards ("You") and Nature Publishing Group ("Nature Publishing Group") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Nature Publishing Group, and the payment terms and conditions.

**All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.**

|                                        |                                                                                       |
|----------------------------------------|---------------------------------------------------------------------------------------|
| License Number                         | 3343591077241                                                                         |
| License date                           | Mar 07, 2014                                                                          |
| Licensed content publisher             | Nature Publishing Group                                                               |
| Licensed content publication           | Nature Reviews Cancer                                                                 |
| Licensed content title                 | How nucleotide excision repair protects against cancer                                |
| Licensed content author                | Errol C. Friedberg                                                                    |
| Licensed content date                  | Oct 1, 2001                                                                           |
| Volume number                          | 1                                                                                     |
| Issue number                           | 1                                                                                     |
| Type of Use                            | reuse in a dissertation / thesis                                                      |
| Requestor type                         | academic/educational                                                                  |
| Format                                 | print and electronic                                                                  |
| Portion                                | figures/tables/illustrations                                                          |
| Number of figures/tables/illustrations | 1                                                                                     |
| Figures                                | Figure with legend for Figure 2. The Essential Features of Nucleotide Excision Repair |
| Author of this NPG article             | no                                                                                    |
| Your reference number                  |                                                                                       |
| Title of your thesis / dissertation    | Molecular Profiling of Lung Cancer                                                    |
| Expected completion date               | Mar 2014                                                                              |
| Estimated size (number of pages)       | 240                                                                                   |
| Total                                  | 0.00 GBP                                                                              |
| Terms and Conditions                   |                                                                                       |

**Terms and Conditions for Permissions**

Nature Publishing Group hereby grants you a non-exclusive license to reproduce this material for this purpose, and for no other use, subject to the conditions below:

1. NPG warrants that it has, to the best of its knowledge, the rights to license reuse of this material. However, you should ensure that the material you are requesting is original to Nature Publishing Group and does not carry the copyright of another entity (as credited in

the published version). If the credit line on any part of the material you have requested indicates that it was reprinted or adapted by NPG with permission from another source, then you should also seek permission from that source to reuse the material.

2. Permission granted free of charge for material in print is also usually granted for any electronic version of that work, provided that the material is incidental to the work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version. Where print permission has been granted for a fee, separate permission must be obtained for any additional, electronic re-use (unless, as in the case of a full paper, this has already been accounted for during your initial request in the calculation of a print run). NB: In all cases, web-based use of full-text articles must be authorized separately through the 'Use on a Web Site' option when requesting permission.
3. Permission granted for a first edition does not apply to second and subsequent editions and for editions in other languages (except for signatories to the STM Permissions Guidelines, or where the first edition permission was granted for free).
4. Nature Publishing Group's permission must be acknowledged next to the figure, table or abstract in print. In electronic form, this acknowledgement must be visible at the same time as the figure/table/abstract, and must be hyperlinked to the journal's homepage.
5. The credit line should read:  
Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)  
For AOP papers, the credit line should read:  
Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)

**Note: For republication from the *British Journal of Cancer*, the following credit lines apply.**

Reprinted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)  
For AOP papers, the credit line should read:  
Reprinted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)

6. Adaptations of single figures do not require NPG approval. However, the adaptation should be credited as follows:

Adapted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)

**Note: For adaptation from the *British Journal of Cancer*, the following credit line applies.**

Adapted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)

7. Translations of 401 words up to a whole article require NPG approval. Please visit <http://www.macmillanmedicalcommunications.com> for more information. Translations of up to a 400 words do not require NPG approval. The translation should be credited as follows:

Translated by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication).

**Note: For translation from the *British Journal of Cancer*, the following credit line applies.**

Translated by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)

We are certain that all parties will benefit from this agreement and wish you the best in the use of this material. Thank you.

Special Terms:

v1.1

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK501244674.

Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

**Make Payment To:**  
Copyright Clearance Center  
Dept 001  
P.O. Box 843006  
Boston, MA 02284-3006

For suggestions or comments regarding this order, contact RightsLink Customer Support: [customercare@copyright.com](mailto:customercare@copyright.com) or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing \$0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

---

---

**Case: 03354038 "Permission to reproduce an image from a PlosOn [ ref:\_00DU0Ifis.\_500U0BFZfr:ref ]**

noreply@salesforce.com on behalf of plosone [plosone@plos.org]

**Sent:** 10 March 2014 19:05

**To:** Richards, Elizabeth

Dear Elizabeth Richards,

Thank you for contacting PLOS ONE.

All content of articles published in PLOS journals is open access. You can read about our open access license here: <http://www.plos.org/about/open-access/>. To summarize, this license allows you to download, reuse, reprint, modify, distribute, and/or copy articles or images in PLOS journals, so long as the original creators are credited (e.g., including the article's citation and/or the image credit).

There are many ways to access our content, including HTML, XML, and PDF versions of each article. Higher resolution versions of figures can be downloaded directly from the article. Additionally, our articles are archived at PubMed Central (<http://www.pubmedcentral.gov/>), which offers a public FTP service from which you can download a complete site of files for each of our articles. See <http://www.pubmedcentral.gov/about/openftlist.html> for links to the FTP site, file list, and organizational directory.

Kind Regards,

Anna Collier  
PLOS ONE  
Staff EO

Case Number: 03354038  
ref:\_00DU0Ifis.\_500U0BFZfr:ref

----- Original Message -----

From: Richards, Elizabeth [elizabeth.richards08@imperial.ac.uk]  
Sent: 07/03/2014  
To: plosone@plos.org  
Subject: Permission to reproduce an image from a PlosOne article in a PhD thesis

Hi

I would like to check that open access licence for Plos One articles (<http://www.plosone.org/static/license.js?sessionId=0ED523A947E6ED421129A3D516F80C64>) covers figures as well as text content.

If you could confirm this as soon as possible, I would be most grateful.

Yours sincerely,

Elizabeth Richards

**NATURE PUBLISHING GROUP LICENSE  
TERMS AND CONDITIONS**

Mar 20, 2014

This is a License Agreement between Elizabeth J Richards ("You") and Nature Publishing Group ("Nature Publishing Group") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Nature Publishing Group, and the payment terms and conditions.

**All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.**

|                                        |                                                       |
|----------------------------------------|-------------------------------------------------------|
| License Number                         | 3343630543091                                         |
| License date                           | Mar 07, 2014                                          |
| Licensed content publisher             | Nature Publishing Group                               |
| Licensed content publication           | Nature Reviews Drug Discovery                         |
| Licensed content title                 | Cellular processing of platinum anticancer drugs      |
| Licensed content author                | Dong Wang and Stephen J. Lippard                      |
| Licensed content date                  | Mar 24, 2005                                          |
| Volume number                          | 4                                                     |
| Issue number                           | 4                                                     |
| Type of Use                            | reuse in a dissertation / thesis                      |
| Requestor type                         | academic/educational                                  |
| Format                                 | print and electronic                                  |
| Portion                                | figures/tables/illustrations                          |
| Number of figures/tables/illustrations | 1                                                     |
| High-res required                      | no                                                    |
| Figures                                | Figure 3   Formation and effects of cisplatin adducts |
| Author of this NPG article             | no                                                    |
| Your reference number                  |                                                       |
| Title of your thesis / dissertation    | Molecular Profiling of Lung Cancer                    |
| Expected completion date               | Mar 2014                                              |
| Estimated size (number of pages)       | 240                                                   |
| Total                                  | 0.00 GBP                                              |
| Terms and Conditions                   |                                                       |

**Terms and Conditions for Permissions**

Nature Publishing Group hereby grants you a non-exclusive license to reproduce this material for this purpose, and for no other use, subject to the conditions below:

1. NPG warrants that it has, to the best of its knowledge, the rights to license reuse of this material. However, you should ensure that the material you are requesting is original to Nature Publishing Group and does not carry the copyright of another entity (as credited in

the published version). If the credit line on any part of the material you have requested indicates that it was reprinted or adapted by NPG with permission from another source, then you should also seek permission from that source to reuse the material.

2. Permission granted free of charge for material in print is also usually granted for any electronic version of that work, provided that the material is incidental to the work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version. Where print permission has been granted for a fee, separate permission must be obtained for any additional, electronic re-use (unless, as in the case of a full paper, this has already been accounted for during your initial request in the calculation of a print run). NB: In all cases, web-based use of full-text articles must be authorized separately through the 'Use on a Web Site' option when requesting permission.
3. Permission granted for a first edition does not apply to second and subsequent editions and for editions in other languages (except for signatories to the STM Permissions Guidelines, or where the first edition permission was granted for free).
4. Nature Publishing Group's permission must be acknowledged next to the figure, table or abstract in print. In electronic form, this acknowledgement must be visible at the same time as the figure/table/abstract, and must be hyperlinked to the journal's homepage.
5. The credit line should read:  
Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)  
For AOP papers, the credit line should read:  
Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)

**Note: For republication from the *British Journal of Cancer*, the following credit lines apply.**

Reprinted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)  
For AOP papers, the credit line should read:  
Reprinted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)

6. Adaptations of single figures do not require NPG approval. However, the adaptation should be credited as follows:

Adapted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)

**Note: For adaptation from the *British Journal of Cancer*, the following credit line applies.**

Adapted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)

7. Translations of 401 words up to a whole article require NPG approval. Please visit <http://www.macmillanmedicalcommunications.com> for more information. Translations of up to a 400 words do not require NPG approval. The translation should be credited as follows:

Translated by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication).

**Note: For translation from the *British Journal of Cancer*, the following credit line applies.**

Translated by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)

We are certain that all parties will benefit from this agreement and wish you the best in the use of this material. Thank you.

Special Terms:

v1.1

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNK501244757.

Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

**Make Payment To:**  
**Copyright Clearance Center**  
**Dept 001**  
**P.O. Box 843006**  
**Boston, MA 02284-3006**

For suggestions or comments regarding this order, contact RightsLink Customer Support: [customer care@copyright.com](mailto:customer care@copyright.com) or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing \$0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

---

---