Identification and Analysis of Differentially Expressed Genes in Human Pulmonary Adenocarcinoma

This thesis has been submitted for the degree of Doctor of Philosophy at the University of Leicester

by

Debra Brocksmith
BSc, MBChB

October 1997

Centre for the Mechanisms of Human Toxicity & Department of Pathology
University of Leicester
Leicester
Abstract

Debra Brocksmith

*The identification and analysis of differentially expressed genes in human pulmonary adenocarcinoma*

Increasing evidence indicates a shift is occurring in the distribution of lung cancer subtypes, with relative and absolute increases in the number of adenocarcinomas. The exact reason for this remains unclear. Molecular and cytogenetic studies have identified several genetic changes occurring in the development of pulmonary adenocarcinoma but many more remain unknown. Until recently, identification of unknown genes relied heavily on the use of linkage-based studies, which are both laborious and time-consuming. Recently, a new technique has been developed to investigate gene expression with greater accuracy and experimental ease: differential display reverse transcription polymerase chain reaction (DDRT-PCR).

In this study DDRT-PCR has been applied to search for novel, differentially expressed genes in human pulmonary adenocarcinoma, using both human normal and adenocarcinoma cell lines and matched surgical tissues. Human cell lines were used in preliminary experiments to adapt and establish the experimental techniques. Results from these studies produced three differentially expressed bands, one of which was a true positive. This cDNA tag was found to show 100% sequence homology to human fibronectin mRNA. Fibronectin is known to show different patterns of expression in various tumours. Similar studies were set up using human surgical specimens collected from two major thoracic units. Results from one matched human tissue sample identified eight differentially expressed bands. Again, only one of these demonstrated true differential expression with a northern blot.

Expression of the tissue-positive band, called H8, was diminished in pulmonary adenocarcinoma RNA samples and may represent a novel tumour suppressor gene. The sequence data that has been obtained from the H8 cDNA tag appears to correspond to the 3’ untranslated region. Northern blotting studies revealed high levels of H8 in normal lung tissue matched to other types of lung tumour, normal liver, stomach and colon, with lower expression in the corresponding tumour samples.
Dedication

This thesis is dedicated to anyone who ever thought they had a good idea!
Acknowledgements

Firstly, I would like to thank Drs. Maggie Manson and Jacqui Shaw for all their help and advice over the past 3 years. They both possess the two qualities necessary for any PhD supervisor; the unfailing ability to remain cheerful in the face of apparent disaster and an aptitude for creative interpretation of experimental data! (The phrases “did you add enzyme?” and “it’s not that bad... really!” will stay with me for years to come!) Thanks also go to Dr. Victor Oreffo for staying calm when I wasn’t, and for listening patiently every time I moaned about RNA/PCR/life!

May I also thank Dr. Rosemary Walker, Histopathology Dept., Glenfield Hospital, Leicester, Dr. Chris Edwards, Histopathology Dept, Birmingham Heartlands Hospital, and the staff & patients of the Thoracic Surgery Departments at both hospitals for their help in providing the tissue samples. Thank you also to Dr. Howard Pringle for his advice on in situ hybridisation.

I would like to acknowledge the following gifts:

K-ras primers from Dr. Victor Oreffo, MRC Unit, Leicester.
Cell line A-549 from Dr. Anna Clouter, MRC Unit, Leicester.
Cell line MOR-P from Ms. Karen Wright, MRC Unit, Cambridge.
Mitochondrial ISH probe from Dr. Howard Pringle, Pathology Dept., Leicester.

A BIG thank you to the Medical Research Council for giving me this marvellous opportunity and for generously funding the project in its entirety.

I would also like to thank everyone in Lab 313 and the all staff at the Unit (especially Fiona, AiLi, David, Lucinda & Alka) for putting up with my eccentricities, making sure I had lunch and reminding me what day it was!

Finally, I would like to thank my family and friends for their continuing support and encouragement, for pointing out the funny side of it, and for helping me to keep it all in perspective.
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>1</td>
</tr>
<tr>
<td>Dedication</td>
<td>2</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>3</td>
</tr>
<tr>
<td>Contents</td>
<td>4-9</td>
</tr>
<tr>
<td>List of figures</td>
<td>10-12</td>
</tr>
<tr>
<td>List of tables</td>
<td>13</td>
</tr>
<tr>
<td>Abbreviations</td>
<td>14-17</td>
</tr>
<tr>
<td>A first thought...</td>
<td>18</td>
</tr>
<tr>
<td><strong>Chapter 1 Introduction to pulmonary adenocarcinoma</strong></td>
<td>19-75</td>
</tr>
<tr>
<td>1.1 Epidemiology of pulmonary adenocarcinoma</td>
<td>20-24</td>
</tr>
<tr>
<td>1.1.1 General changes in lung cancer incidence</td>
<td>20</td>
</tr>
<tr>
<td>1.1.2 Changes in the incidence of pulmonary adenocarcinoma</td>
<td>21</td>
</tr>
<tr>
<td>1.2 Aetiology of pulmonary adenocarcinoma</td>
<td>25-38</td>
</tr>
<tr>
<td>1.2.1 Active tobacco smoking</td>
<td>25</td>
</tr>
<tr>
<td>1.2.2 Passive tobacco smoking</td>
<td>27</td>
</tr>
<tr>
<td>1.2.3 Environmental and occupational agents</td>
<td>28</td>
</tr>
<tr>
<td>1.2.4 Genetic susceptibility</td>
<td>30</td>
</tr>
<tr>
<td>1.2.5 Sex differences</td>
<td>33</td>
</tr>
<tr>
<td>1.2.6 Dietary factors</td>
<td>37</td>
</tr>
<tr>
<td>1.2.7 Viral agents</td>
<td>38</td>
</tr>
<tr>
<td>1.3 Molecular genetics of pulmonary adenocarcinoma</td>
<td>39-50</td>
</tr>
<tr>
<td>1.3.1 Oncogene activation</td>
<td>39</td>
</tr>
<tr>
<td>1.3.2 Tumour suppressor gene mutation</td>
<td>45</td>
</tr>
<tr>
<td>1.3.3 Growth factors and growth factor receptors</td>
<td>48</td>
</tr>
<tr>
<td>1.3.4 Anti-metastasis and other genes</td>
<td>49</td>
</tr>
<tr>
<td>1.4 Pathology of pulmonary adenocarcinoma</td>
<td>50-60</td>
</tr>
<tr>
<td>1.4.1 Normal lung histology</td>
<td>50</td>
</tr>
<tr>
<td>1.4.2 Classification of lung cancer</td>
<td>53</td>
</tr>
<tr>
<td>1.4.3 Diagnostic criteria for histological subtypes</td>
<td>53</td>
</tr>
</tbody>
</table>
3.2.2 Extraction of RNA using modified acid GITC protocol

3.2.3 RNA purification

3.2.4 RNA quality control

3.2.5 mRNA extraction

3.3 Precautions for differential display RT-PCR

3.3.1 Treatment of equipment

3.3.2 Treatment of reagents

3.3.3 Negative controls for DDRT-PCR

3.4 DDRT-PCR

3.4.1 cDNA synthesis by reverse transcription

3.4.1 cDNA amplification by PCR

3.4.3 Separation and display of PCR products

3.5 Extraction and re-amplification of bands

3.5.1 Extraction of bands from sequencing gel

3.5.2 Re-amplification of cDNA from bands

3.6 Subcloning of template DNA

3.6.1 Ligation

3.6.2 Transformation and screening

3.6.3 Plasmid preparation and purification of DNA

3.7 Sequencing of template DNA

3.7.1 Manual sequencing of template DNA

3.7.2 Automated sequencing of template DNA

3.7.3 M13 single-stranded sequencing

3.8 Blotting of RNA and genomic DNA

3.8.1 RNA dot blots

3.8.2 RNA northern blots

3.8.3 Southern blot of genomic DNA

3.9 Preparation of cDNA probes and hybridisation

3.9.1 cDNA probe preparation

3.9.2 Hybridisation conditions

3.9.3 Stripping of probe from membrane

3.10 Solid phase RT-PCR using magnetic beads
3.11 RACE-PCR 118-119
3.11.1 First strand synthesis 118
3.11.2 Adaptor ligation 118
3.11.3 5' RACE-PCR 119
3.12 In situ hybridisation 120-123
3.12.1 Oligonucleotide probe design 120
3.12.2 Oligonucleotide probe labelling 121
3.12.3 Pre-treatment of slides and hybridisation 121
3.12.4 Post-hybridisation washes and detection 122
3.12.5 Controls for ISH 123
3.13 Computer software 124-125

Chapter 4 Adaptation and development of the DDRT-PCR technique for use with human lung RNA 126-149

4.1 Strategy for development of the DDRT-PCR method for use with human lung RNA 127
4.2 Results and trouble-shooting experiments for DDRT-PCR using human lung RNA 130
4.2.1 Investigation of "ghost" lanes in DDRT-PCR gels 130
4.2.2 Attempts to increase differential display band signal 133
4.2.2.1 Improving RNA extraction and quantification 133
4.2.2.2 Improving labelling efficiency for DDRT-PCR products 135
4.2.2.3 Increasing PCR efficiency 135
4.2.3 Investigation of persistent lung tissue RNA degradation 139
4.2.3.1 Comparison of RNA extraction methods 141
4.2.3.2 Comparison of RNA extracted from other tissues 144
4.2.3.3 Investigation of RNase contamination of reagents 144
4.3 Discussion 147-149
### Chapter 5 Differential mRNA expression between a normal and a lung tumour cell line

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 Introduction</td>
<td>150</td>
</tr>
<tr>
<td>5.2 Results</td>
<td>151</td>
</tr>
<tr>
<td>5.2.1 DDRT-PCR</td>
<td>151</td>
</tr>
<tr>
<td>5.2.2 Cloning and sequencing of the differentially displayed bands</td>
<td>151</td>
</tr>
<tr>
<td>5.2.3 Northern and Southern blotting and probe hybridisation</td>
<td>153</td>
</tr>
<tr>
<td>5.2.4 Further analysis of FN sequence tag</td>
<td>159</td>
</tr>
<tr>
<td>5.2.4.1 Sequencing of FN cDNA tag</td>
<td>159</td>
</tr>
<tr>
<td>5.2.4.2 Expression studies using FN probe</td>
<td>161</td>
</tr>
<tr>
<td>5.3 Discussion</td>
<td>165-170</td>
</tr>
<tr>
<td>5.3.1 Explanation of experimental results</td>
<td>165</td>
</tr>
<tr>
<td>5.3.1.1 Behaviour of MOR-P cell line with FN probe</td>
<td>165</td>
</tr>
<tr>
<td>5.3.1.2 GBP &amp; c72 results</td>
<td>165</td>
</tr>
<tr>
<td>5.3.1.3 FN expression studies</td>
<td>166</td>
</tr>
<tr>
<td>5.3.2 Significance of increased FN expression in tumour cell line</td>
<td>166</td>
</tr>
</tbody>
</table>

### Chapter 6 Differential mRNA expression between matched human lung tumour samples

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1 Introduction</td>
<td>171</td>
</tr>
<tr>
<td>6.2 Results</td>
<td>172</td>
</tr>
<tr>
<td>6.2.1 DDRT-PCR</td>
<td>174</td>
</tr>
<tr>
<td>6.2.2 Cloning and sequencing of the differentially displayed bands</td>
<td>174</td>
</tr>
<tr>
<td>6.2.3 Northern and Southern blotting and probe hybridisation</td>
<td>176</td>
</tr>
<tr>
<td>6.2.3.1 Preliminary investigations of differential expression</td>
<td>179</td>
</tr>
<tr>
<td>6.2.3.1.1 Probe competition assay with H3, H5 &amp; H6</td>
<td>179</td>
</tr>
<tr>
<td>6.2.3.3 Densitometry analysis of H8 sequence tag</td>
<td>187</td>
</tr>
<tr>
<td>6.3 Discussion</td>
<td>190-195</td>
</tr>
<tr>
<td>6.3.1 Explanation of experimental results</td>
<td>190</td>
</tr>
<tr>
<td>6.3.1.1 Problems of 3'-UTR</td>
<td>190</td>
</tr>
<tr>
<td>6.3.1.2 Why the use of T12MT was discontinued</td>
<td>193</td>
</tr>
<tr>
<td>6.3.2 Why use human lung tissue?</td>
<td>193</td>
</tr>
</tbody>
</table>
Chapter 7  Further analysis of the cDNA band H8

7.1  Introduction

7.2  Results

7.2.1  Full sequence analysis of H8 cDNA tag

7.2.1.1  Manual sequencing results

7.2.1.2  Automated cycle sequencing results

7.2.1.3  Sequencing using single-stranded M13 cloning products

7.2.1.4  Database identification of H8 cDNA tag

7.2.1.5  Peptide map of H8 sequence

7.2.2  Further expression studies using H8 cDNA

7.2.2.1  Northern blotting results

7.2.2.2  Qualitative RT-PCR

7.2.3  5' RACE-PCR

7.2.4  In situ hybridisation

7.3  Discussion

7.4  General discussion

7.5  Conclusions

References

Appendix I - Materials

Appendix II - Primer sequences

Appendix III - Buffers, media & other reagents

A final thought...

“The unofficial rules of DDRT-PCR !”
List of figures

Chapter 1 Introduction to pulmonary adenocarcinoma

1.1 Lung cancer mortality trends in England and Wales 22
1.2 Mechanism of ras activation 40
1.3 Normal human airway histology 52
1.4 Photomicrographs of main types of human lung cancer 58-59

Chapter 2 Introduction to DDRT-PCR

2.1 Principles of differential display RT-PCR 79
2.2 Scheme of work for DDRT-PCR 80
2.3 Schematic diagram of principles of suppression subtractive hybridisation facing 87

Chapter 3 Methods

3.1 Map of cloning vector pCR™ II 101
3.2 Dideoxynucleotide sequencing reaction 105
3.3 Single-stranded sequencing using M13 phage 108
3.4 Principles of solid phase RT-PCR using magnetic beads 117

Chapter 4 Adaptation and development of the DDRT-PCR technique for use with human lung RNA

4.1 DDRT-PCR using control RNA samples with primer API facing 128
4.2 Summary of action for DDRT-PCR trouble-shooting experiments 131
4.3 Ghost lanes on DDRT-PCR gel facing 132
4.4 DDRT-PCR gel demonstrating a lack of improvement in signal from the normal lanes when using increased RNA template in the first strand RT step facing 136
4.5 DDRT-PCR gel demonstrating a lack of improvement in signal from the normal lanes when using increased cDNA template in the PCR step facing 137

4.6 Selected lanes from a DDRT-PCR gel comparing the signal obtained from the use of $^{32}$P and $^{35}$S labelled nucleotide facing 138

4.7 Mg$^{2+}$/dNTP titration results facing 140

4.8 Comparison of RNAzol™ B and acid GITC RNA extraction protocols facing 142

4.9 RT-PCR with K-ras primers using RNA from cases 1-5 facing 143

4.10 RT-PCR & DDRT-PCR results using RNA extracted from matched non-pulmonary human tissue samples facing 145

4.11 Degradation in lung tissue RNA samples facing 146

Chapter 5 Differential mRNA expression between a normal and a lung tumour cell line

5.1 DDRT-PCR gel autoradiograph from cell line RNA study facing 152

5.2 Examples of Genbank database sequence comparisons of cloned cell line cDNA fragments 154

5.3 a/b Cell line sample blots hybridising to FN and GAPDH probes facing 155

5.3 c Hybridisation studies using GBP2 cDNA probe facing 156

5.3 d Hybridisation studies using c72720 cDNA probe facing 157

5.4 Hybridisation of FN probe to northern and Southern blots facing 160

5.5 txt Human fibronectin cDNA sequence (3' end) 161

5.6 Human tissue RNA samples hybridising with FN probe facing 162

5.7 Mouse RNA samples hybridising with FN probe facing 163

5.8 Structure of the fibronectin molecule 168

Chapter 6 Differential mRNA expression between matched human lung tumour samples

6.1 DDRT-PCR gel autoradiograph from tissue study facing 175

6.2 Partial sequence data on human tissue cDNA clones 177-178
6.3 Flow chart showing analysis of human tissue RNA DDRT-PCR sequence tags 180
6.4 a Dot blot hybridisation results using cDNA probes H2, 3, 5, 6 and 7 facing 181
6.4 b Hybridisation results with H1 and H4 cDNA probes facing 182
6.4 c Example of diffuse smearing with H1 and H4 cDNA probes facing 183
6.5 a Northern blot hybridisation with H8 and GAPDH probes using cell line RNA samples facing 185
6.5 b Southern blot hybridisation with H8 and GAPDH probes using cell line RNA samples facing 186
6.6 Sequence data for H3 and H6 showing the presence of repeat elements 188
6.7 Strategies used to obtain additional 5' sequences from a DDRT-PCR cDNA tag 191-192

Chapter 7 Further analysis of the H8 cDNA tag

7.1 Preliminary sequencing results for H8 template DNA 198
7.2 txt Restriction enzyme digest fragments from H8 insert 202
7.3 PCR analysis of recombinant bacterial clones facing 204
7.4 Manual sequencing data from H8 restriction fragments cloned into M13 mp18 vector 205
7.5 Complete H8 cDNA tag sequence 206
7.6 Northern blot hybridisation with matched lung tumour RNA facing 210
7.7 Human tissue RNA samples hybridising with H8 cDNA probe facing 212
7.8 PCR primers designed from H8 sequence 214
7.9 a/b Results of RT-PCR using H8 primers facing 215
7.9 c Results of RT-PCR using H8 primers facing 216

Key: txt = figure lies within page of text
List of tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Summary of histological time trends in lung cancer</td>
<td>24</td>
</tr>
<tr>
<td>1.2</td>
<td>Revised WHO histological classification of lung tumours</td>
<td>54-55</td>
</tr>
<tr>
<td>1.3</td>
<td>Staging of lung cancer</td>
<td>63-64</td>
</tr>
<tr>
<td>1.4</td>
<td>Prognosis of lung cancer cell types</td>
<td>66</td>
</tr>
<tr>
<td>3.1</td>
<td>Summary of human tissue samples collected</td>
<td>91</td>
</tr>
<tr>
<td>3.2</td>
<td>Computer software programs used</td>
<td>124</td>
</tr>
<tr>
<td>4.1</td>
<td>Human tissue RNA sample OD readings</td>
<td>134</td>
</tr>
<tr>
<td>5.1</td>
<td>Summary of hybridisation results from DDRT cell line probes</td>
<td>158</td>
</tr>
<tr>
<td>5.2</td>
<td>Summary of northern blotting results using human tissue</td>
<td>164</td>
</tr>
<tr>
<td>6.1</td>
<td>Matched pulmonary adenocarcinoma cases collected</td>
<td>172</td>
</tr>
<tr>
<td>6.2</td>
<td>Details of bands obtained from human case #1</td>
<td>174</td>
</tr>
<tr>
<td>6.3</td>
<td>Database sequence homologies of H-band partial cDNA tags</td>
<td>176</td>
</tr>
<tr>
<td>6.4</td>
<td>Summary of hybridisation results from DDRT-PCR human tissue probes</td>
<td>184</td>
</tr>
<tr>
<td>6.5</td>
<td>Densitometry readings for H8 and GAPDH probes</td>
<td>189</td>
</tr>
<tr>
<td>7.1</td>
<td>Results of H8 template variation with cycle sequencing reaction</td>
<td>200</td>
</tr>
<tr>
<td>7.2</td>
<td>Database sequence homologies of H8 cDNA tag</td>
<td>207</td>
</tr>
<tr>
<td>7.3</td>
<td>H8 amino acid sequence using G(C)G PEPDATA program</td>
<td>208</td>
</tr>
<tr>
<td>7.4</td>
<td>Summary of northern blotting results using human tissue</td>
<td>211</td>
</tr>
<tr>
<td>7.5</td>
<td>Anti-sense H8 oligo sequences used for in situ hybridisation</td>
<td>218</td>
</tr>
</tbody>
</table>
Abbreviations

AAH atypical adenomatous hyperplasia
AC adenocarcinoma
AHH aryl hydrocarbon hydroxylase
AP arbitrary primer
APC adenomatous polyposis coli gene
ASO allele specific oligonucleotide
B(a)P benzo(a)pyrene
BAC bronchioloalveolar cell carcinoma
BCIP 5’ bromo-4’ chloro-3’indoyl phosphate
BSA bovine serum albumin
Ca cancer (tumour)
CAMP cyclophosphamide/ doxorubicin/ methotrexate/ cisplatin
CAP cyclophosphamide/ doxorubicin/ cisplatin
cDNA complementary DNA
COAD chronic obstructive airways disease
CoCl₂ cobalt chloride
CRC Cancer Research Campaign
CsCl caesium chloride
CYP cytochrome p450 mono-oxygenase
DD differential display
ddNTP dideoxynucleotide triphosphate
DEPC diethyl pyrocarbonate
dH₂O distilled water
DNA deoxyribonucleic acid
DNase deoxyribonuclease
dNTP deoxynucleotide triphosphate
ds double-stranded
DTT dithiothreitol
DXT deep X-ray treatment (radiotherapy)
EC European Community
EDTA  ethylene diaminetetracetate
EGF(R)  epidermal growth factor (receptor)
ER  estrogen receptor
ETS  environmental tobacco smoke
FGF  fibroblast growth factor
FB  flexible bronchoscope
FHIT  fragile histidine triad gene
FN  fibronectin
GAP  GTPase activating protein
GAPDH  glyceraldehyde 3-phosphate dehydrogenase
GBP  guanylate binding protein
GITC  guanadinium thiocyanate
GNRP  guanine nucleotide releasing protein
GSP  gene-specific primer
GST  glutathione-S-transferase
GTE  glucose/tris/EDTA
HPV  human papilloma virus
HRT  hormone replacement therapy
IMS  industrial methylated spirits
IPTG  isopropyl-β-D-thiogalactopyranoside
ISH  in situ hybridisation
IVC  inferior vena cava
KA  potassium acetate/acetic acid
kb  kilobase
KOAc  potassium acetate
LB  Luria-Bertani
LCC  large cell cancer (lung)
LiCl  lithium chloride
LOH  loss of heterozygosity
M-MLV  Moloney murine leukaemia virus
MCC  mutated in colonic cancer gene
MgCl₂  magnesium chloride
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>MHC</td>
<td>major histocompatibility complex</td>
</tr>
<tr>
<td>MOPS</td>
<td>3-(N-morpholino)propane sulphonic acid</td>
</tr>
<tr>
<td>MPC</td>
<td>magnetic particle concentrator</td>
</tr>
<tr>
<td>MRC</td>
<td>Medical Research Council</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>MTS</td>
<td>multiple tumour suppressor gene</td>
</tr>
<tr>
<td>N</td>
<td>normal (non-tumour)</td>
</tr>
<tr>
<td>NAC</td>
<td>N-acetyl cysteine</td>
</tr>
<tr>
<td>NaCl</td>
<td>sodium chloride</td>
</tr>
<tr>
<td>NaOAc</td>
<td>sodium acetate</td>
</tr>
<tr>
<td>NaOH</td>
<td>sodium hydroxide</td>
</tr>
<tr>
<td>NBT</td>
<td>nitroblue tetrazolium</td>
</tr>
<tr>
<td>NNK</td>
<td>4-methylnitrosamine-1-(3)-pyridyl-1-butanone</td>
</tr>
<tr>
<td>NSCLC</td>
<td>non-small cell lung cancer</td>
</tr>
<tr>
<td>OD</td>
<td>optical density</td>
</tr>
<tr>
<td>PAC</td>
<td>P1 artificial chromosome</td>
</tr>
<tr>
<td>PAH</td>
<td>polycyclic aromatic hydrocarbon</td>
</tr>
<tr>
<td>PAS</td>
<td>pulmonary adenoma susceptibility</td>
</tr>
<tr>
<td>PBS</td>
<td>phosphate buffered saline</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PDGF</td>
<td>platelet derived growth factor</td>
</tr>
<tr>
<td>PEG</td>
<td>polyethylene glycol</td>
</tr>
<tr>
<td>PIREMA</td>
<td>primer introduced restriction with enrichment for mutant alleles</td>
</tr>
<tr>
<td>PK</td>
<td>proteinase K</td>
</tr>
<tr>
<td>PTP</td>
<td>protein tyrosine phosphatase gene</td>
</tr>
<tr>
<td>RACE</td>
<td>rapid amplification of cDNA ends</td>
</tr>
<tr>
<td>RAP</td>
<td>random arbitrary priming</td>
</tr>
<tr>
<td>RAR</td>
<td>retinoic acid receptor gene</td>
</tr>
<tr>
<td>RDA</td>
<td>representational difference analysis</td>
</tr>
<tr>
<td>RF</td>
<td>replicative form</td>
</tr>
<tr>
<td>RI</td>
<td>recombinant inbred</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
</tbody>
</table>
RNase  ribonuclease
RPMI  Roswell Park Memorial Institute
RT  reverse transcriptase/transcription
SCLC  small cell lung cancer
SDS  sodium dodecyl-sulphate
SEER  surveillance/epidemiology/end results
SPA  sheep pulmonary adenomatosis
ss  single-stranded
SSC  standard saline citrate
SSH  suppression subtractive hybridisation
SURF  selective UV radiation fractionation
T<sub>12</sub>MN  poly(T) anchor primer
Taq  *Thermus aquaticus*
TBE  tris-borate-EDTA
TBS  tris buffered saline
TE  tris-EDTA
TEMED  N,N,N',N'-tetramethylenediamine
TGF  transforming growth factor
TNM  tumour/nodes/metastases
Tris  tris(hydroxymethyl)amino methane
TSG  tumour suppressor gene
TSNA  tobacco-specific N-nitrosamine
UTR  untranslated region
UV  ultraviolet
*VHL*  Von Hippel Lindau gene
WHO  World Health Organisation
X-gal  5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside
"There is nearly a complete consensus of opinion that primary malignant neoplasms of the lung are among the rarest forms of disease."

Adler, 1912
Chapter 1 - Introduction to pulmonary adenocarcinoma
1.1 Epidemiology of pulmonary adenocarcinoma

1.1.1 General changes in lung cancer incidence

Lung cancer is the most common form of malignant tumour in the world today: it accounts for nearly 18% of male cancers world-wide and for 22% of male cancers in developed countries [Parkin et al, 1993]. In 1985 there were approximately 900,000 new cases of lung cancer world-wide with more than 60% of these occurring in developed countries [Parkin et al, 1993]. This figure represents a 36% increase since 1980 and current opinion suggests the incidence will continue to rise for the foreseeable future. Within the European Community (EC) lung cancer is also the most prevalent form of neoplasm. It remains the leading cause of male cancer with 135,000 new cases per year, while in women it is the seventh most common form of cancer, accounting for 23,000 new cases annually [CRC, 1992]. In the United Kingdom (UK) it is again the most widespread malignancy representing 1 in 6 of all new cases of cancer [CRC, 1992].

Unlike other diseases the epidemiology of lung cancer is discussed mainly in terms of its mortality rate rather than its incidence rate. This is because, unlike other diseases, the incidence of lung cancer is virtually equal to the mortality. In the UK, lung cancer is responsible for 25% of all cancer mortality and for 6% of all mortality. It is the commonest cause of cancer death in men and in women is second only to breast cancer, although in Scotland and some areas of northern England lung cancer mortality is now in excess [CRC, 1992]. In 1990 there were 39,000 deaths from lung cancer in the UK, equivalent to 100 deaths per day and 75% of these occurred in men. In 1990 more than 500,000 people in the developed world died from this disease and it is predicted that this figure will have risen to several million deaths annually by the first decades of the next century [Parkin et al, 1993].

The scale of lung cancer as a health problem has escalated throughout this century. In Europe and the US in the early 1900’s it was a relatively uncommon disease with fewer than 10 deaths per 100,000 men per year. Within 50 years there had been a 600% increase to 60 deaths per 100,000 men per year. Current studies now reveal the mortality from lung cancer to lie at 104 deaths per 100,000 men per year in Europe, 82 deaths per 100,000 men per year in the US and 61 deaths per 100,000 men per year in the UK [Lopez-Abente et al, 1995, Devesa et al, 1995, CRC, 1992]. Lung cancer mortality in women has always been lower than that
observed in men, but rates have increased dramatically since the 1950’s, with deaths per 100,000 women per year rising from <5 in 1950 to >30 in 1990 (US) [Ernster, 1994]. In the US and some parts of Europe it now surpasses breast cancer as the leading cause of female cancer death [Ernster, 1994].

Lung cancer mortality also varies with age, the older groups having the highest mortality rates. For example, the mortality per 100,00 men aged 65-74 lies at 500 compared to just 6 per 100,000 for those men aged 35-44 (UK, 1991). Recently a decrease in male mortality has been observed for all ages but particularly in younger men. Although a similar trend has been reported in women under 55 years, the overlying female mortality is increasing [CRC, 1992]. See figure 1.1 for an overview of these trends.

1.1.2 Changes in the incidence of pulmonary adenocarcinoma

Lung cancer is a heterogeneous group of tumours that is divided into two main types: non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). The former, which accounts for 75% of all cases is further divided into squamous cell carcinoma (SCC), adenocarcinoma (AC) and large cell carcinoma (LCC). SCLC is a distinct clinicopathological entity derived from neuroendocrine cells and forms the remaining 25% of cases. Historically, SCC has always been the most prevalent type of lung cancer, particularly in male cigarette smokers, accounting for nearly 50% of all lung cancer cases. Adenocarcinoma has shown a stronger association with females and non-smokers and has accounted for 20% of all lung cancer cases. It has been noted since the mid-1970’s that a shift is occurring in the distribution of lung cancer cell types throughout the world, with an increase in the relative and absolute numbers of AC being reported in both sexes. Numerous studies in various parts of the world reveal an increase of AC without any similar increase in SCC. Data obtained from RPMI study 1962-1975 [Vincent et al, 1977], SEER study 1974-1984 [Dodds et al, 1986] SEER study 1973-1987 [Travis et al, 1995] and US hospitals study 1977-1984 [Wynder and Covey , 1987] all reported increases in the incidence of AC, particularly in men, while the incidence of SCC fell. A study examining cases from the mid-south US found that in 1964 AC accounted for 13% and SCC for 44% lung cancer cases in women, but by 1985 AC had increased to 31% and SCC had fallen to 32% of such cases [El-Torky et al, 1990]. Figures for male lung cancer changed from 4% AC and 50 % SCC in 1964 to 27% AC and 37% SCC in 1985. The SEER
Figure 1.1 a) Lung cancer mortality trends England & Wales 1940-1990

![Graph showing lung cancer mortality trends from 1940 to 1990 in England & Wales. The graph includes a key indicating mortality rates for males and females in different age groups (35-44 yrs, 45-54 yrs, 55-64 yrs, 65-74 yrs). The years 41 to 91 are shown on the x-axis, and the rates per 10^5 pop. are shown on the y-axis.]

Key: Males ■ ■  Females ■ ■
35-44 yrs — — — — 45-54 yrs — — — — — —
55-64 yrs — — — — — — 65-74 yrs — — — — — —

Figure 1.1 b) Lung cancer as a percentage of all cancer deaths (UK 1990)

<table>
<thead>
<tr>
<th></th>
<th>Males</th>
<th>All people</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td>32%</td>
<td>26,924 deaths</td>
<td>39,269 deaths</td>
<td>12,345 deaths</td>
</tr>
</tbody>
</table>

Redrawn from CRC Factsheet 11.1, 1992 (see References).
study 1973-1987 found AC had replaced SCC as the predominant tumour type, with AC representing 31.5% and SCC 29.4% of all lung cancer cases [Travis et al, 1995]. Tong et al [1996] found AC accounting for 34% and SCC representing only 24% of lung cancer cases in a survey from 1986-1993. This trend is not confined to the US with studies from elsewhere revealing similar findings. Bourke et al [1992] looked at lung cancer in younger patients from US, Italy and Israel and found AC was the leading cell type in patients aged <45 years (56-69% cases). He also demonstrated AC occurred frequently in patients >45 years, 27-44% cases AC compared to 30-39% cases SCC. Oriental countries have traditionally had a slightly different frequency distribution for lung tumour types, with AC being the predominant cell-type over SCC, particularly in women. In Japan 1970-1989, the incidence of AC was seen to rise from 26% to 45% in males and from 45% to 66% in females. For the same time period, SCC fell from 48% to 33% in males and from 24% to 17% in females [Ikeda et al, 1991]. Similar changes have been reported in Hong Kong, which has one of the highest lung cancer mortality rates particularly for women. A study examining cases from 1960-1990 found an increase in AC incidence in both sexes, with male AC rising from 16% to 26% and female AC rising from 34% to 59%. Again, SCC incidence fell in both sexes from 44% to 39% in men and from 23% to 16% in women [Lam et al, 1993]. See table 1.1.

While no-one can argue that a change in the pattern of lung cancer cell types has indeed been occurring over the past 20 years, the relevance of this observation has been questioned. Some researchers feel the data are simply due to changes in diagnostic and classification protocols, while others believe the trends are real and are due to changes in etiological factors. Various hypotheses have been suggested to explain the rise in AC, among them changes in smoking habits, dietary factors, environmental toxins and viruses (see section 1.2).
Table 1.1. Summary of histological time trends in lung cancer.

<table>
<thead>
<tr>
<th></th>
<th>1964</th>
<th>1985</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Male AC</td>
<td>4%</td>
<td>27%</td>
<td>[El-Torky, 1990]</td>
<td></td>
</tr>
<tr>
<td>Male SCC</td>
<td>50%</td>
<td>37%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female AC</td>
<td>13%</td>
<td>31%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female SCC</td>
<td>44%</td>
<td>32%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male AC</td>
<td>26%</td>
<td>45%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male SCC</td>
<td>48%</td>
<td>33%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female AC</td>
<td>45%</td>
<td>66%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female SCC</td>
<td>24%</td>
<td>17%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-US &gt; 45 yrs old</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC frequency</td>
<td>27-44%</td>
<td></td>
<td>[Bourke, 1992]</td>
<td></td>
</tr>
<tr>
<td>SCC frequency</td>
<td>30-39%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hong Kong (1960-1990)</td>
<td>1960</td>
<td>1990</td>
<td>[Lam, 1993]</td>
<td></td>
</tr>
<tr>
<td>Male AC</td>
<td>16%</td>
<td>26%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male SCC</td>
<td>44%</td>
<td>39%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female AC</td>
<td>34%</td>
<td>59%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female SCC</td>
<td>23%</td>
<td>16%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC</td>
<td>-</td>
<td>31.5%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCC</td>
<td>-</td>
<td>29.4%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC</td>
<td>-</td>
<td>34%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCC</td>
<td>-</td>
<td>24%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Key:** % = AC or SCC as a proportion of all cases of lung cancer  
AC = pulmonary adenocarcinoma  
SCC = squamous cell carcinoma of lung

This table shows the differences in the % of AC at various time points from different studies and compares this to the change in SCC cases. These studies demonstrate an increase in % AC and a decrease in % SCC.
1.2. Aetiology of pulmonary adenocarcinoma

1.2.1 Active tobacco smoking

A cigarette smoker has a 15-20 times greater risk of developing lung cancer than a non-smoker [Wald and Hackshaw, 1996] and the overwhelming majority of lung cancer cases (>80-90%) are caused by smoking tobacco [CRC, 1992]. In Northern Europe and USA, smoking is responsible for more than 90% of male and 70-80% of female lung cancers [Parkin et al, 1992]. Eighty five percent of all new lung cancer cases in men and 48% of such cases in women are also attributed to tobacco smoking [Parkin et al, 1993]. The prevalence of smoking tobacco increased steadily throughout the first half of this century. The figures for male cigarette consumption then plateaued until the mid-1970’s and began to fall. It was not until after World War 2 that smoking became a common habit in women and, as seen with men, cigarette consumption also peaked in the 1970’s. Unlike men, however, the fall in female smoking has not been as rapid and for the first time the percentage of female smokers (35%) is now greater than that of male smokers (30%) [CRC, 1992]. The type and composition of cigarettes has also changed over the years, with a sharp increase in the proportion of smokers using lower tar and filtered cigarettes. It is now estimated that 1 in 3 of the global population over the age of 15 years smokes tobacco, making the total number of current smokers in excess of 1 billion people in the world today [Chapman, 1996]. Some $6 \times 10^{12}$ cigarettes are currently consumed per year, causing an estimated 3 million deaths annually [Chapman, 1996] and this figure is expected to rise to approximately 10 million deaths per year by 2020. Given that lung cancer occurs in at least 15% of smokers [Law, 1990], the number of deaths due to this disease will be in excess of 1.5 million per year in the early decades of the next century - a depressing statistic with which to begin the new millennium.

Historically, lung cancers have been divided into 2 groups in relation to cigarette smoking; the so-called Kreyburg group I tumours in which there was a strong causal relationship and Kreyburg group II tumours in which the link was much weaker [Doll and Hill, 1964, Kreyburg, 1969]. The former group included SCC and SCLC, while AC fell into the latter. A study by Jedrychowski et al [1992] demonstrated the relative risk for smoking as 13.5 for SCC and 15.4 for SCLC, while for AC it was only 3.1. The study also observed a similar pattern in the attributable risk for smoking with figures of 90% for SCC, 80% for
SCLC and only 64% for AC. This would imply that AC is more likely to be related to factors other than tobacco smoking than either SCC or SCLC. Jedrychowski, however, also found a linear dose-response relationship between daily cigarette consumption and all types of lung cancer, including AC. An earlier study by Weiss *et al* [1972] gave similar results, challenging the accepted view that group 2 tumours were not related to smoking. The literature remains divided, but many groups have provided data indicating cigarette smoking is a weaker risk factor for AC [Harris *et al*, 1993, Shimizu *et al*, 1994, Muscat and Wynder, 1995, Tong *et al*, 1996].

How then could tobacco smoke be causally related to the development of AC? The combustion of tobacco produces a smoke consisting of 95% gaseous compounds and 5% particulate compounds. The total particulate matter, also called the tar, contains known initiators, promoters and co-carcinogens. Two groups of compounds are of special interest in lung tumour pathogenesis: polycyclic aromatic hydrocarbons (PAHs) and tobacco-specific N-nitrosamines (TSNAs). Activated PAHs are produced from the metabolic activation of procarcinogens by cytochrome p450 enzymes, while the TSNAs are derived from nicotine and are formed during both tobacco processing and smoking. PAHs are considered to act as contact carcinogens and have been shown to induce the development of SCC, but not AC, on intratracheal installation in rodents [Blacklock, 1957, Deutsch-Wenzel *et al*, 1983]. However, one particular PAH, benzo(a)pyrene (BaP), has been used to induce lung adenocarcinomas in mice by the production of point mutations in K-ras, predominantly at codon 12 [You *et al*, 1989]. This oncogene mutation is commonly observed in pulmonary AC (for further details see section 1.3). TSNAs are also powerful animal carcinogens and the concentration of these compounds in tobacco products is high. For example, NNK is present in mainstream smoke in 0.12-3.7 μg per cigarette [Hoffmann and Hoffmann, 1994]. Two types of TSNAs, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) and 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol, both induce mainly adenocarcinoma in the lungs of several rodent species [Hecht, 1989, Rivenson *et al*, 1988]. In male smokers, AC accounts for 30% of lung cancers, making it the second most frequent lung cancer sub-type. In female smokers, AC is the most common form of lung cancer, representing more than 40% of lung tumours [Wynder and Covey, 1987, El-Torky *et al*, 1990, Muscat and Wynder, 1995]. If the tar content and nicotine yields have been reduced and the use of filtered cigarettes has increased to nearly 100%, why are we still seeing more lung cancers, particularly AC, in smokers? Ironically, it may be
Chapter 1  Introduction to pulmonary adenocarcinoma

precisely because of these factors that the number of AC has risen. The composition of cigarettes has been altered to produce less smoke by making the tobacco undergo a more complete combustion. This has been achieved not only by designing more efficient filters but also by using tobacco blends containing higher nitrate levels. This has resulted in decreased amounts of PAHs and increased amounts of TSNAs, particularly NNK [Hoffmann and Hoffmann, 1994]. In addition, smokers are taking deeper inhalations when using low nicotine cigarettes and are holding the smoke in the airways for longer before exhaling. Thus, forcing the particulate matter deeper into the peripheral airways and exposing the lining cells to the contents for a longer time [Heming et al, 1981, Kozlowski et al, 1982]. It is in these peripheral airway sites that the adenocarcinomas are found, and it is not difficult to imagine how exposure to the constituents of tobacco smoke could induce AC development in humans as observed in rodents. With greater numbers of people consuming these nitrate-rich cigarettes, perhaps the evolution of increased numbers of AC and decreased numbers of SCC can be at least partially explained.

1.2.2 Passive tobacco smoking

Although the vast majority of lung cancers occur in smokers, an estimated 15% of cases occur in non-smokers [Law, 1990]. Adenocarcinomas are the most common type of lung cancer reported in non-smokers, accounting for between 50-80% of lung tumours in this group [Kabat and Wynder, 1984, Jedrychowski et al, 1992, Muscat and Wynder, 1995]. What factors are responsible for the development of AC in these people? Together with radon, tobacco smoke is the most prevalent indoor air pollutant occurring in industrialised countries. Studies examining tobacco exposure in non-smokers found that up to 75% of such individuals demonstrated detectable levels of markers for tobacco smoke [Patel, 1995]. It is currently estimated that 25% of lung cancers in non-smokers are directly attributable to environmental tobacco smoke (ETS) inhalation, otherwise known as passive smoking [CRC, 1992]. Passive smokers have >1.5x greater risk of developing lung cancer than true non-smokers [Law, 1990]. Studies examining passive smoking and lung cancer development have found the proportion of AC ranging from 43-100% [Wu et al, 1995, Brownson et al, 1992]. A large proportion of these passive smokers are women and it has been hypothesised that other factors, such as hormonal and reproductive status, may be implicated in the development of
their lung cancers. Several problems have been apparent in studies trying to determine the exact extent of ETS as a risk factor; firstly, the dose of carcinogens received through ETS exposure will be much lower than that of mainstream smoke inhalation and consequently tests which measure dose-response data must be of suitable sensitivity, and secondly, much of the epidemiological data obtained from self-reporting of exposure levels has been found to be unreliable [Nilsson, 1996]. Many studies have used detection of cotinine, a nicotine metabolite, to estimate smoking status. However, this compound has a short half-life and is known to be produced in differing amounts according to individual variations in nicotine metabolism. As a result, some heavy smokers who have recently given up may appear to be life-long non-smokers and some current smokers may not be detected at all. A further problem of some of these studies is the possible confounding by dietary factors such as saturated fat intake, which may be linked with deficiency of antioxidant dietary elements. All of these points serve to hinder what is already a difficult and controversial research issue. As the saying goes “there is no smoke without fire”, and in the case of AC, at least some of this smoke is from tobacco combustion.

1.2.3 Environmental & occupational agents

It is estimated that approximately 8% of all lung cancers are due to occupational exposure to carcinogens. Only bladder cancer has a higher proportion of such cases [Steeland et al, 1996]. Several occupational and environmental agents have been implicated in the pathogenesis of lung cancer and particularly AC. They include the following; radon, asbestos, arsenic, heavy metals, PAHs and wood dust.

Radon (\(^{222}\text{Rn}\)) is a radioactive gas produced when uranium (\(^{238}\text{U}\)) decays to unstable lead. These radon “daughters” are metal ions that bind to particles in the air and are deposited in the respiratory tree on inhalation. They then damage the surrounding tissue by irradiation with \(\alpha\)-particles, which causes chromosomal breaks. The synergistic effect of smoking has complicated risk assessment in some studies. However, an increase in all types of lung cancer has been demonstrated in uranium miners, regardless of smoking history. This is primarily due to a relative increase in SCLC [Churg, 1994, Patel, 1995], but increases in both SCC and AC have also been reported [Archer et al, 1974]. Although the concentration of radon occurring in domestic homes is much lower than that found in mines, the prolonged period of time people
spend in their homes may result in a significant overall exposure. Mortality data from miners show a dose-response pattern even at low cumulative radon exposures and suggest there is no absolute risk threshold [Steenland et al, 1996].

Asbestos is the generic name given to a number of naturally occurring inorganic fibrous silicates. The shape and dimensions of each group of asbestos fibres confer differences in disease potential, with the long thin fibres of the crocidolite (blue asbestos) group causing the majority of asbestos-related neoplasms. Occupational exposure occurs mainly in construction and maintenance workers where asbestos has been used as insulation material and in brake linings. Asbestos exposure has been claimed to cause increased AC in the lung, but unfortunately many studies are flawed in their selection of suitable control groups with regard to age and smoking history. It now appears that there is only an increased risk of lung cancer if asbestosis is present. This is the term used to denote pulmonary interstitial fibrosis caused by the inhalation of asbestos. Reports not taking this into account are merely recording smoking-induced tumours [Hughes and Weill, 1991]. The association of asbestos exposure and lung tumour cell types is also confounded by the synergism with smoking and as most of the current literature is derived from studies with smokers, insufficient evidence is available to link it to any one sub-type at present [Churg, 1994].

Arsenic exposure occurs primarily in smelting workers and those exposed to arsenic-based pesticides. Unlike other occupational agents, there is strong data from human epidemiological studies of its carcinogenicity but very little evidence from animal experiments [Steenland et al, 1996]. Arsenic acts by interfering with DNA repair and has been shown to cause chromosomal abnormalities. As with asbestos studies, the literature gives variable results with different groups reporting increases in the incidence of each of the major lung cancer types [Churg, 1994]. These studies also suffer from the lack of good control group selection regarding smoking history and age. Studies monitoring exposure to other heavy metals such as chromium, nickel and cadmium, have also provided similar results.

Polycyclic aromatic hydrocarbons are not only present in cigarette smoke but are also found in diesel engine exhaust fumes and in the fumes from various cooking oils. The effect of diesel fumes has been difficult to study due to the problems of documenting past exposure. Most studies have used non-specific measurements such as nitrous oxides to represent levels of these emissions. Increasingly specific tests are being developed and current studies will provide more accurate information on the exact risk levels posed by these agents [Steenland et
al, 1996]. The risks associated with the heating of edible oils are well known and may cause the high rates of AC observed in oriental women, despite their lower incidence of tobacco smoking. Chinese cooking pots heat edible oils to more than 265 °C, thus generating large quantities of fumes. These fumes are known to contain carcinogens such as dibenzo(a,h)anthracene and B(a)P, it is likely that many women inhale the fumes for several hours on a daily basis and after a prolonged exposure this may lead to the development of AC [Shuguang, 1994].

Finally, wood dust exposure is well established in the causation of AC in the sinuses and nasal passages. It is likely that the dust also induces such tumours in the airways, as the mucosa has a similar composition to that observed in the nose and sinuses [Patel, 1995].

1.2.4 Genetic susceptibility

Although environmental factors play a crucial role in the aetiology of lung cancers, the fact that not all those exposed to these agents develop the disease would suggest that host factors in tumour susceptibility are also important. These factors include differences in metabolic enzyme activity, DNA repair, oncogene expression and general genetic instability.

Several studies have found increased risk of lung cancer in the relatives of lung cancer patients. Estimates of risk elevation range from x 2 in smokers to x 4 in non-smokers [Economou et al, 1994]. It has also been documented that lung cancer patients who survived a first tumour have subsequently developed a second malignancy, either in the lung or at another site [Economou et al, 1994]. This may reflect the effect of smoking or may result from a predisposing genetic abnormality. Studies linking blood group with lung cancer predisposition have shown mixed results, as have similar studies examining the role of HLA antigens [Economou et al, 1994]. Such studies operate without an association-hypothesis, and if enough variables are compared some will show statistical significance purely by chance. No studies indicating an excess risk of AC have yet been demonstrated.

As discussed above, many environmental agents are lung carcinogens but it is their metabolites, rather than the parent compounds, that are implicated in pulmonary tumourigenesis. The enzymes that metabolise chemical carcinogens are divided into 2 groups: phase I and phase II enzymes. The phase I enzymes are cytochrome p450 mono-oxygenases
(CYP) which catalyse the production of reactive oxygenated intermediates. These are then conjugated to form water-soluble compounds by the phase II enzymes before being excreted. Although these enzymes are primarily expressed in the liver, they are also found in many extra-hepatic organs. Many studies have attempted to determine the effects of metabolic enzyme genotypes on lung cancer susceptibility. Of the phase I enzymes, CYP1A1, CYP2D6 and CYP2E1 have been most consistently implicated in lung cancer susceptibility. CYP1A1 is involved in the metabolism of PAH compounds such as B(a)P via its aryl hydrocarbon hydroxylase (AHH) activity [Raunio et al, 1996]. AHH activity was noted to be increased in smoking-induced lung cancers. However, studies have shown conflicting results and it now appears that AHH activity is a poor susceptibility marker [Economou et al, 1994, Raunio et al, 1996]. A polymorphism at the Msp I restriction site in CYP1A1 has been reported, showing increased lung cancer in homozygotes of the rare allele m2 in the Japanese population [Kawaraji et al, 1990]. This polymorphism shows a marked ethnic variation and is rare (<2% homozygous) in Caucasian populations. CYP2D6, associated with debrisoquine hydroxylase metabolism, shows polymorphisms leading to 2 phenotypes in humans; the autosomal recessive "poor metaboliser" (PM) phenotype, and the heterozygous/ homozygous dominant "extensive metaboliser" (EM) phenotype. The EM phenotype is over-represented among lung cancer patients, particularly SCC and SCLC [Economou et al, 1994]. Interestingly, it is least associated with AC, the type of lung cancer showing the weakest association with smoking. These findings suggest two things. Firstly, CYP2D6 may activate one of the chemicals in tobacco smoke and this is supported by its ability to activate NNK. Secondly, the CYP2D6 locus is physically close to the true causative gene and we are merely observing the effects of linkage [Economou et al, 1994, Raunio et al, 1996]. CYP2E1 metabolises several known carcinogens, including TSNAs. Restriction enzyme polymorphisms of the gene for this enzyme have been associated with lung cancer in Oriental groups but, as seen with CYP1A1, translating the significance of these findings to other ethnic populations has been difficult [Kato et al, 1995]. A recent study by El-Zein et al [1997] reported a significant association between CYP2E1 and the occurrence of pulmonary AC. It is interesting to note that some of these enzymes are found in specific parts of the lung. For example, CYP1A1 is located in the periphery within the cell types that are likely to give rise to AC [Anittila et al, 1991]. It is possible that chemical carcinogens from tobacco smoke act either as substrates for these
enzymes or serve to induce their expression. This may then result in the development of a tumour.

The activity of the phase II conjugation enzymes also appears to influence lung cancer susceptibility, the most well known of these is glutathione-S-transferase (GST). The GSTM1 gene encodes the isoenzyme GST Mu, the substrates of which include several epoxides from PAH metabolism. This locus contains 3 alleles, GSTM1-A and B and the deficient GSTM1 null allele. Up to 50% of Caucasians are known to be homozygous for the null allele and are devoid of any GSTM1 activity. Preliminary epidemiological studies appear to indicate that the null genotype confers an excess risk of developing some types of lung cancer [Sun et al, 1997]. These results are by no means clear-cut and several potential confounding factors have yet to be eliminated [Economou et al, 1994, Raunio et al, 1996]. The other main group of phase II enzymes are the N-acetyltransferases. The population is divided into “slow” and “fast” acetylators. Several studies have linked slow acetylation of arylamines with bladder cancer but as yet there is little data on lung cancer susceptibility [Raunio et al, 1996].

An individual’s susceptibility to the effect of chemical carcinogens is likely to depend upon the combination of both phase I and II enzyme activities. Several reports have examined phase I-II combination effects, the results are variable. CYP1A1 Msp I mutation combined with GSTM1 null genotype was found to be associated with x 6-7 increased risk of smoking-associated lung cancers (i.e. not AC) in Japanese populations [Kawajiri et al, 1995]. These individuals also had higher levels of $p53$ mutation. Smokers with CYP1A1 Msp I polymorphism have been reported to show x 5 elevated levels of K-ras and $p53$ mutation frequencies than those with non-susceptible genotypes [Kawajiri et al, 1996] and when combined with the GSTM1 null genotype, mutation frequencies increased to x 11 for K-ras and x 9 for $p53$. The association of increased $p53$ abnormality has also been observed with GSTM1 null genotype in Mediterranean groups, occurring this time in AC and SCLC [To-Figueras et al, 1996]. These synergistic increases in mutation frequency in patients with combinations of susceptible genotypes support the idea that there is increased PAH adduct formation, most likely from cigarette smoking. Another possible lung cancer susceptibility factor is from a deficiency in DNA repair. Comparatively less work has been done on the role of DNA repair mechanisms in lung cancer susceptibility. Initial studies indicate that some lung cancer patients show deficient excision repair compared with controls [Wei et al, 1996].
It must be remembered that the metabolism of many compounds depends upon the action of a number of enzymes and is, therefore, controlled by a number of genes, the proportional effect of each one varying tremendously. Also, the activity of each enzyme may have its own environmental and genetic determinants. Thus it would be naive to oversimplify these processes and assume that only one or two are significant or to examine them in isolation. Many of the differences identified show ethnic specificity in terms of gene structure and allelic distribution, making results difficult to extrapolate between populations. None of the genetic markers studied have been clearly linked to increased lung cancer susceptibility, although the most consistent appears to be the combination of CYP1A1 allele m2 with GSTM1 null genotypes. The problems of assessing an individual's background tumour susceptibility are compounded by several factors. These include the exposure of different populations to many different chemical agents, the high prevalence of respiratory disease/infection and a widespread exposure to passive smoking. Additional factors, such as genetic heterogeneity, epistatic interaction and the low frequency of familial case-clusters relative to the frequency of the disease in the general population, make the establishment of a polygenic model for human lung cancer extremely difficult [Dragani et al., 1996]. It has also been argued that measuring familial case-clusters of a disease is a relatively insensitive method of indicating genetic predisposition to that disease. Cancers having significant genetic components may show undetectable familial clustering [Law, 1990]. Genetic susceptibility is only a measure of the likelihood of an individual's risk of developing cancer, patients with cancer may not be genetically susceptible and neither may their relatives. Matched controls from the so-called "normal" population will also be a mixture of susceptible and non-susceptible individuals. By comparing these groups in the absence of other disease markers, any effects will be diluted out and result in a smaller estimate of excess risk than if the comparison had been purely between genetically susceptible and non-susceptible groups. Thus, the size of any genetically determined inter-individual variation in lung cancer risk remains unknown.

1.2.5 Sex differences

Since the mid-20th century, female lung cancer mortality has increased by more than 500% [Ernster, 1996, Zang and Wynder, 1996] while lung cancer mortality in males has
levelled off. In 1950, lung cancer accounted for only 3% of cancer mortality in women, by
1995 this had increased to 24% [Ernster, 1996]. Over the same time period, the number of
women smokers has increased and female exposure to tobacco carcinogens is now
approaching that observed in men [CRC, 1992]. It would not be unreasonable, therefore, to
expect a shift in the distribution of lung cancer in women to those cell types associated with
male lung cancer, particularly SCC. This has not been the case, however, and the proportion
of female AC has risen instead. The incidence of pulmonary AC is far higher in women than
men, a long-standing enigma of lung cancer research. In the US, AC accounts for up to 37%
of female and 25% of male lung cancers [Dodds et al, 1996] and in Hong Kong the proportion
of AC is 55% in women and 34% in men [Lung et al, 1992]. When the figures for non­
smoking women are examined the proportion of AC cases becomes even greater, rising as high
as 65-77% in both Caucasian and Oriental populations [Gao et al, 1987, Osann, 1991,

What is it about females that makes them more likely to develop AC than their male
counterparts and why should so many female non-smokers develop this form of lung cancer?
Several epidemiological studies suggest that women may be more sensitive to the effects of
tobacco carcinogens than men. Levels of DNA-adducts were higher in female lung tissue than
male adduct levels when adjusted for equivalent smoking doses [Ryberg et al, 1994, Kure et
al, 1996]. Higher adduct levels were also associated with shorter duration of smoking and
lower smoking dose before tumour onset [Ryberg et al, 1994]. High levels of stable adducts in
a target tissue appear to be directly related to the probability of tumour formation within that
tissue [Swenberg et al, 1985] and their occurrence with lower exposure levels would support
the hypothesis of increased female susceptibility. Female lung cancers were also found to
contain higher frequencies of p53 mutation G:C — » T:A transversions than males (40% female
v. 25% male) [Guinee et al, 1995], presumably resulting from DNA-adduct formation. It has
been estimated that women have between 1.5 -3.0 times greater risk of developing smoking­
related lung tumours than men [Ryberg et al, 1994, Guinee et al, 1995]. However, only about
45% female AC are associated with cigarette use [Osann, 1991], the vast majority are found
occurring in non-smokers which suggests that other factors may be implicated in female AC
pathogenesis.

Alavanja et al [1995] stated that 48% of lung cancers in current non-smokers and 31%
in lifetime non-smokers were due a combination of 5 main risk factors; non-malignant lung
Chapter 1  Introduction to pulmonary adenocarcinoma
disease, environmental tobacco exposure, previous history of smoking, family history of lung
cancer and saturated fat intake. Previous history of chronic obstructive airways disease
(COAD) such as bronchitis and emphysema, and of respiratory infections such as pneumonia
are associated with lung cancer independently of smoking. Impaired pulmonary function could
allow increases in the levels of pulmonary carcinogens either by reduced clearance or by
potentiation of their effects as a result of inflammatory processes. Wu et al [1995] found a
significant increase in risk of developing AC in non-smoking women with a history of either
asthma or COAD. Similar results have been observed in other studies [Gao et al, 1987,
Tockman, 1994]. Wu et al [1996] also found x 4 increased risk of developing AC in non-
smoking women associated with family history of lung cancer. This could merely reflect
environmental tobacco exposure, however. Over the past decade the scientific plausibility of
environmental tobacco smoke causing lung cancer in large numbers of individuals has gained
credence. Fontham et al [1994] found that the relative risk of ETS exposure was equivalent
for all lung cancer cell types, even though the numbers of AC (76%) among non-smoking
females was far greater than for the other types combined (24%). Studies measuring cotinine
levels as a marker of tobacco smoke exposure have demonstrated that the relationship between
AC and ETS in women is much weaker than for other lung cancer cell types [Ellard et al,
1995, De Waard et al, 1995]. Studies of exposure to occupational risk factors have been
mostly carried out on male subjects and relatively little is known about female exposure.
suggested exposure to environmental carcinogens from cooking oil fumes was responsible for
the excess of AC cases observed there. Gao’s group also made an interesting observation that
hormonal factors may be involved as AC risk seemed to be associated with ovarian activity.
Several studies examining the role of sex hormones and lung cancer have been undertaken.
Short menstrual cycles, late menopause and the use of estrogen replacement therapy have all
been linked with increased risk of AC development [Gao et al, 1987, Taoili and Wynder,
1994]. Estrogen receptors (ERs) have been demonstrated in NSCLC tissues [Chaudhuri et al,
1982, Canver et al, 1994] and have been found to occur with greater frequency in female lung
cancers [Kaiser et al, 1996]. Kaiser’s group found that 69% of AC in women demonstrated
functional ERs compared with 42% in male AC. Other non-pulmonary types of
adenocarcinoma also demonstrate varying frequencies of ER expression, but no differences in
activity between males and females has yet been established [Theodoropoulos et al, 1995].
35
There are several biological mechanisms that may help to explain the possible pathogenesis of female pulmonary AC. Firstly, there may be gender-specific differences in metabolic or DNA repair enzyme activity that would account for the increased adduct levels associated with the handling of PAH-compounds, resulting in greater DNA damage in susceptible females [Ryberg et al, 1994]. Animal data has demonstrated sex differences in the activity of p450 enzymes in rats, with liver and lung p450s from males showing increased binding capacity and subsequent xenobiotic oxidation ability [Kato, 1974]. In humans, it has been shown that the levels of p450 enzymes in hair follicles from men and women differ, with men showing higher levels [Sawaya, 1994]. This could result in diminished metabolic activation and subsequent detoxification of lung carcinogens in women. Thus, giving support to the hypothesis that women may be inherently more susceptible to tobacco carcinogens than men. Secondly, the effect of sex hormones on lung tumour development may be significant. Kato [1974] stated that sex hormones may alter the level of p450s in rat microsomes, with androgens increasing p450 content by 20-30%. Estrogens may also directly affect tumour differentiation, as guinea pigs injected with estradiol went on to develop pulmonary adenomas, while those given simultaneous doses of other steroids were protected from estrogen-induced tumourigenesis [Lupulescu, 1983]. Interestingly, while there is little evidence to suggest the oral contraceptive pill has any association with female AC, estrogen hormone replacement therapy (HRT) does appear to be linked to an increased risk of AC in post-menopausal women, particularly in association with smoking [Taioli and Wynder, 1994]. The implications of these findings are of importance because of the increased numbers of female smokers and also because of the increased frequency with which HRT is now prescribed. It has also recently been shown that estradiol activates the c-erbB-2 receptor [Matsuda et al, 1993]. This oncogene is frequently amplified or over-expressed in AC of the breast and lung. Immunohistochemical staining has demonstrated the frequency of c-erbB-2 receptors in female lung cancers is twice that seen in male tumours [Guinee et al, 1995], and a similar finding was reported with the ratio of p53 mutations. The p53 mutations may occur as a result of direct/indirect exposure to tobacco smoke via active/passive smoking. If the loss of normal p53 were to result in the amplification of the c-erbB-2 oncogene, any lung tumours bearing this mutation would have an inherent growth advantage in females due to the high levels of circulating estrogens. It may transpire that estrogens activate other oncogene receptors that may also result in the growth of AC.
1.2.6 Dietary factors

Dietary components are believed to be of great significance in the pathogenesis of many tumours. There have been two principle dietary correlations in lung cancer; an inverse association with fruit and vegetable intake, and a direct association with the consumption of saturated fats and cholesterol. Fruit and vegetables are rich in vitamin A precursors, the provitamin A carotenoids, and it was generally accepted that β-carotene in particular was largely responsible for this protection, presumably via its anti-oxidant properties. The protective effect of fruit and vegetables appears to be greater in men than in women and is also stronger for heavy smokers than non-smokers [Stryker et al, 1988, Zeigler et al, 1991]. However, one study found the strongest protective effect was observed in female non-smokers with AC and LCC [Koo et al, 1988]. The results of a large, randomised trial in Finland which examined the effect of vitamin supplementation and lung cancer in male smokers, found no such protective effects from β-carotene [ATBC study, 1994] and similar results have been found by others [Zeigler et al, 1996]. Against expectations, cases given β-carotene supplements showed a higher incidence of lung cancer and had a excess mortality of 8%. It is now felt that the case selection was flawed in that only heavily active smokers were used, most of whom continued to smoke throughout the trial. The significance of vitamins in lung cancer pathogenesis remains somewhat controversial.

Other studies have reported elevated risk of lung cancer with high fat intake. Different groups report conflicting results. One study quotes dietary cholesterol to be significantly greater in cases of SCC than AC [Goodman et al, 1988], while another claims the association of dietary cholesterol and lung cancer is linked purely to AC [Jain et al, 1990]. A study by Alavanja et al [1993] found an 11-fold increase in the risk of AC in non-smoking women with a high intake of saturated fat. The association between high-fat diet and increased incidence of several types of cancer is well known but the causal mechanism has not yet been established [Ames et al, 1995]. Whether the risk is solely due to a constituent within the food or is due to the way in which the food is prepared and cooked remains unknown.
1.2.7 Viral agents

AC is divided into several sub-types, one of which is brochioloalveolar cell carcinoma (BAC). BAC has increased dramatically over the past decade and some workers attribute the increase in AC to this observation [Barsky et al., 1994]. BAC has several distinguishing features including high incidence in non-smokers and females and a tendency to be multifocal. BAC also shares certain histological characteristics with ovine pulmonary carcinoma, also called sheep pulmonary adenomatosis (SPA) or jaagsiekte. SPA is also multifocal and is believed to be caused by a retrovirus, the jaagsiekte retrovirus [Palmarini et al., 1996]. Although no evidence of viral involvement in the human disease has been demonstrated, it raises the possibility that a viral agent, known or unknown, may be linked with AC pathogenesis.

A second possible viral association with AC of the lung comes from a more familiar agent, HIV. The occurrence of lung cancer in patients with HIV/AIDS is not common, but those that do develop these tumours exhibit a higher proportion of AC than other cell types. AC accounts for between 50-60% of lung cancers in HIV/AIDS patients [Fraire and Awe, 1995, Vyzula and Remick, 1996] and the reason for this is not clear. It may be that the biological features of HIV-lung cancer are different due to immune deficiency, or that AC develop because of the increased incidence of pulmonary pathology associated with AIDS. It may be related to the extreme youth of these patients, or it may just be an artifactual association due to the small numbers involved.
1.3. Molecular genetics of pulmonary adenocarcinoma

Molecular and cytogenetic studies have identified numerous genetic changes occurring in AC and in other forms of lung cancer, the significance of which in terms of tumour initiation and progression are as yet unknown. These changes can alter the regulation of oncogenes, tumour suppressor genes, growth factors and anti-metastasis genes.

1.3.1 Oncogene activation

Oncogenes, or strictly speaking proto-oncogenes, encode proteins that function as either growth factors, growth factor receptors, signal transduction proteins or nuclear proteins involved in the regulation of transcription. When activated by point mutation, amplification, translocation or by the insertion of certain viral promoters, these genes lead to deregulated cellular proliferation. A number of oncogenes have been found to be abnormally expressed in lung cancer, such as \textit{myb}, \textit{raf} and \textit{jun} in SCLC, and \textit{myc} and \textit{ras} in NSCLC [Johnson, 1995]. Only the latter group is of relevance in the pathogenesis of AC, the remainder will not be discussed here.

The \textit{ras} genes (H-ras, K-ras and N-ras) encode a membrane associated protein (p21) possessing GTPase activity, which is involved in signal transduction. When p21 binds to GTP the signal is "switched on" and "switched off" by GTP hydrolysis. The intrinsic rate of GTP hydrolysis and GDP release are low and are increased by regulatory proteins such as GTPase activating protein (GAP) and guanine nucleotide releasing proteins (GNRPs) [Hesketh, 1994]. Point mutations, commonly seen at codons 12, 13 and 61, alter the p21 product sufficiently to prevent binding to GAP, thus keeping p21 in its GTP-bound, activated form. See figure 1.2. Mutant \textit{ras}, therefore, produces transforming ability by the autonomous stimulation of a growth signal. \textit{Ras} mutations are observed in 30% of NSCLC cell lines and over 20% of NSCLC tumours [Rodenhuis \textit{et al}, 1993]. Very few such abnormalities have been detected in SCLC cell lines or tumours. H-ras and N-ras mutations in lung cancer are sporadic and the vast majority of \textit{ras} abnormalities occur in the K-ras gene. Of all lung cancer cell types K-ras mutation is most strongly associated with AC and more than 35% of such cases contain abnormal K-ras alleles. This compares with only 10% of LCC and 5% of SCLC [Rodenhuis, 1993]. More than 85% of these K-ras mutations are due to a G-T transversion at codon 12.
Figure 1.2. Mechanism of ras activation.

a) Normal ras

```
RAS-p21
GTP

"ON"
```

```
RAS-p21
GAP
GTP

"OFF"
```

b) Mutant ras

```
RAS-p21
GTP
GAP

"ON"
```

```
RAS-p21
GDP

No "OFF" signal
```

Point mutation at codon 12, 13 or 61 prevents binding of GAP, thus maintaining p21 in its GTP-bound, activated form. This acts as an autonomous growth signal and may result in the proliferation of a neoplastic clone.
[Rodenhuis, 1988, Slebos et al, 1991]. A recent study using a new detection technique, called PCR-PIREMA (primer introduced restriction with enrichment for mutant alleles) was used by Mills et al [1995] to detect mutant K-ras in bronchoavleolar lavage fluid in patients with lung cancer. A mutation frequency of 56% was reported in the AC group, a figure much higher than those quoted from conventional PCR-ASO studies (allele specific oligonucleotide), due to the increase in sensitivity of this technique. However, this still means that between 50-70% of lung AC do not contain K-ras mutations. Could these cases contain other types of ras activation? Rodenhuis [1987] asked the same question when he looked at expression levels and gene copy number in a series of lung cancer cases. No evidence of overexpression or gene amplification was found. Perhaps the K-ras negative cases of AC contain other mutations which have the same effect on tumour development as K-ras?

Adenocarcinoma of the lung was thus divided into two classes according to the presence or absence of K-ras mutation and workers began to look for differences and similarities between the two groups. The presence or absence of K-ras activation was not associated with differences in tumour histology, age at diagnosis, stage of disease at diagnosis, sex of the patient or the incidence of other tumours. K-ras positive tumours did tend to be less differentiated than those without the mutation, but the reason for this was not clear. Other differences between K-ras positive and negative tumours have revealed some interesting points. Initial studies reached the surprising conclusion that the presence of K-ras activation slowed the growth of the tumour and, thus, was associated with a slightly improved prognosis [Rodenhuis et al, 1988]. However, these preliminary cases had all been relatively early lesions and their numbers were small. Once larger studies were set up it became obvious that this first conclusion was erroneous and that K-ras mutation conferred a strongly unfavourable prognosis, particularly for stage 1 and 2 disease [Rodenhuis and Slebos, 1992, Keohavong et al, 1996, Fukuyama et al, 1997]. A recent study by Visscher et al [1997] also reported tumours with K-ras mutation were more frequently associated with an invasive growth pattern. However, another recent study also examined K-ras activation in advanced stage adenocarcinomas, and found no difference in spread, treatment response or survival between K-ras positive and negative groups [Rodenhuis et al, 1997].

Several large studies to investigate further the significance of K-ras activation in human lung cancer have been undertaken. The largest study looked at 280 cases in total and found K-ras mutation at codon 12 in 30% of AC from patients who smoked, while only 5% of AC from
non-smokers were positive for the mutation [Rodenhuis and Slebos, 1992]. Other studies, both prospective and retrospective, have produced similar results with an overall frequency of 30-35% K-ras positive AC in smokers and of <5% in non-smokers [Silini, 1994, Westra et al, 1993, Husgafvel-Pursianen, 1993]. However, these rates may be found to be greater if the more sensitive PCR-based techniques were applied. SCLC and other classes of NSCLC were not found to contain significant levels of activated mutant ras. This led to the hypothesis that carcinogens within tobacco smoke induce point mutations in K-ras which ultimately lead to the development of AC-type tumours. Animal studies support this, with more than 75% of carcinogen-induced AC containing point mutations at codon 12 of K-ras [Oreffo et al, 1993, You et al, 1989]. Studies from Oriental countries, where pulmonary AC is more common, have revealed a lower incidence of K-ras mutation with an average frequency of 16% in smokers and 3% in non-smokers [Sugio et al, 1992, Lung et al, 1992]. The difference in the frequency of mutant K-ras between Caucasian and Oriental populations may also be due to carcinogen exposure. Either because of differences in tobacco composition, in the type of cigarettes consumed, i.e. use of filters, or exposure of non-tobacco related carcinogens such as those from cooking oil vapour.

The reason for the association between K-ras activation and adenocarcinoma development is not clear at present. Studies examining p21 expression in normal human tissue have demonstrated low p21 levels in normal respiratory epithelium, whereas the levels in secretory epithelium were much greater [Furth et al, 1987]. Another possibility is selective exposure of the smaller airways (from which ACs arise) to carcinogens, possibly the constituents of tobacco smoke [Rodenhuis, 1987]. An alternative explanation is that these tumours are metastases from occult colorectal tumours, in which ras mutations occur with a frequency of 50%. However, it is highly unlikely that so many patients would present in this way and the codon 12 mutation seen in gastrointestinal K-ras positive tumours is a G-A transition rather than the G-T transversion observed in lung cancer [Vogelstein et al, 1988]. It was believed that the specificity of ras genes for particular tumours was due to differences in the protein products. However, no functional differences in the mutant p21s have been found and the functional domain in which supposed effector molecules bind is identical in each p21 protein [Hesketh, 1994]. Work done with transgenic animals has implied that there may not be an absolute specificity between the tumour type and the mutant ras gene, for example mice with H-ras and the albumin enhancer/promoter developed lung AC [Suda, 1987].
In colorectal tumours a high rate of K-ras mutation in non-dysplastic atypical areas suggests it is not the predominant initiating event [Jen et al, 1994]. Observations in animal studies and human colorectal carcinoma also implicate ras activation as an early event, possibly preceding neoplasia. Cells from non-tumourigenic cell lines that have undergone spontaneous transformation are found to possess activating K-ras mutations [Pan et al, 1991]. Mammary tumours induced in neonatal rats contain latent ras mutation in phenotypically normal mammary glands which are then activated by exposure to estrogens [Kumar et al, 1990], and in mouse lung activated K-ras is found in hyperplastic foci [Belinsky et al, 1991]. These observations suggest K-ras mutation is an early event in tumour development. Activated K-ras has been detected in both benign and malignant colorectal lesions and in areas of non-neoplastic regeneration [Chaubert et al, 1994]. In human lung, topographical distribution of ras mutation using SURF (selective UV radiation fractionation) demonstrated K-ras mutation only in malignant cells and with a homogeneous distribution [Li et al, 1994c]. This implies the mutation occurred prior to clonal expansion in a small precursor lesion. No mutation of K-ras was detected in adjacent normal tissue or in areas of hyperplasia, metaplasia or dysplasia [Li et al, 1994c Sugio et al, 1994]. In the mouse lung, K-ras mutation is found in areas of hyperplasia and the nature of the base substitution at codon 12 may determine the malignant progression of the lesion [Nuzum et al, 1990].

What role does K-ras play, if any, in tumorigenesis? Is it necessary for the malignant phenotype or does it merely provide a growth advantage for pre-transformed cells? What are the genetic changes occurring in the K-ras negative tumours? What role is played by ras-related genes? What are the other genetic changes occurring in K-ras positive tumours? What is the importance of carcinogens and tissue-specific susceptibility? At present, the lack of a well-defined pre-malignant respiratory lesion in human AC is preventing these and other questions from being answered. A recent study by Kawano et al [1996] examining NNK-induced murine lung AC, found no evidence of a correlation between tumour proliferation and K-ras mutation. This would imply that the presence of this mutation does not confer a selective growth advantage. Although there is no doubt of a link between K-ras mutation and AC of the lung, questions remain as to its significance. Arguments still persist concerning its temporal location in the molecular chain reaction that culminates in tumour formation. Some studies claim it is an early event [Li et al, 1994c], while others claim K-ras mutations occur late in AC development [Sugio, 1994]. Recently, a third alternative has been proposed in a
report by Westra et al [1996] during the examination of K-ras mutation in atypical adenomatous hyperplasia (AAH) with synchronous AC of the lung. It was demonstrated that many areas of normal lung, AAH and AC contained apparently independent K-ras mutations. This was interpreted as giving further credibility to the concept of field cancerisation being responsible for multifocal lung tumours, and that K-ras mutation may simply be a non-specific change induced by carcinogen exposure. Field cancerisation theory states that multiple cell groups undergo neoplastic transformation under the stress of local carcinogen activity [Slaughter et al, 1953]. Following long-term exposure to tobacco smoke carcinogens, it does not seem unreasonable to assume that pre-neoplastic changes could occur in many cell groups within the respiratory tree. Whether K-ras mutation is irrelevant to subsequent tumour development is a more contentious issue.

Perhaps the development of new detection techniques, such as that reported recently by Keohavong et al [1997], will help reconcile these problems. Keohavong's group have combined a PCR-based mutation allele enrichment assay with denaturing gradient gel electrophoresis to attain a sensitivity of detection of a single mutated cell among $10^5$ normal cells. They demonstrated the presence of mutant K-ras in histologically normal tissue immediately adjacent to K-ras positive lung tumour in 4/8 cases. Whether this is due to the presence of pre-neoplastic lesions or to the diffusion of DNA from necrotic tumour cells is not known. This type of methodology could easily be applied to a topographical study using micro-dissected fixed tissues to help evaluate other suspected mutations involved in lung cancer pathogenesis.

All these data suggest that K-ras mutation in lung cancer is dependent on many factors, including target cell specificity, carcinogen type/dose, tumour stage, metabolic/hormonal factors and genetic susceptibility [Waldmann and Rabes, 1996]. This complicates the use of K-ras as a biomarker for lung cancer development. So, even the significance of the one oncogene that has been associated with pulmonary AC is by no means indisputable and it will be interesting to monitor developments in this area over the next few years.
1.3.2 Tumour suppressor gene mutations

In addition to abnormalities in the sequences and expression of oncogenes, the loss or inactivation of another class of genes may also be important in the development of lung cancer. These are the tumour suppressor genes (TSGs) whose normal function is to suppress cell proliferation. Unlike their oncogene counterparts which require only one allele to mutate to induce transformation, TSGs must have mutations in both alleles. Cytogenetic studies of non-random chromosome loss in tumours and restriction fragment length polymorphism analysis have helped to identify numerous possible TSGs involved in lung cancer pathogenesis. Common cytogenetic abnormalities in lung tumour cells are observed on chromosomes 3p, 13q and 17p, but other sites include 5q, 6p+q, 8p, 9p+q, 11p and 18q. Many of these studies have divided the lung cancer sub-groups only as NSCLC and SCLC, and little information on individual cell types is available at present.

The commonest site for loss of heterozygosity (LOH) in lung cancer is on chromosome 3p, with 100% SCLC and up to 50% NSCLC containing a deletion of this region. LOH refers to the loss of alleles on a chromosome detected by assaying markers for which an individual is constitutively heterozygous. If such a marker binds near the locus of the TSG, it may be lost if part or the whole of the TSG is lost, via deletion, non-disjunction or mitotic recombination events, as part of the mutation which eventually causes tumourigenesis. Further studies have demonstrated at least three distinct regions involved in the 3p deletion; namely 3p25, 3p21 and 3p14. Susceptibility to various other tumours such as renal, uterine, breast and gonadal carcinomas is associated with this region, implying that a significant TSG does exist on chromosome 3p. Several candidate genes have been put forward, including von Hippel Lindau (VHL) [Sekido, 1994], retinoic acid receptor β (RAR-β) [Howle et al, 1993], c-erb-Aβ [Kaye et al, 1990], protein tyrosine phosphatase γ (PTP-γ) [Croce, 1991, LaForgia, 1991], semaphorin (H.sem.IV) [Roche et al, 1996] and, most recently, the fragile histidine triad (FHIT) [Ohta et al, 1996, Sozzi et al, 1996]. Von Hippel Lindau disease is a dominant familial cancer syndrome predisposing to various tumours, including renal cell cancer. The VHL gene maps to 3p25 and is not known to be associated with lung cancer. A recent study showed that although all of 72 lung cancer cell lines examined expressed VHL mRNA, the mutation rate was low with only 5 cell lines demonstrating any aberration [Sekido et al, 1994]. Another candidate TSG is RAR-β, localised to 3p24 and encoding a retinoic acid receptor
protein. The gene is expressed in normal lung tissue and in the majority of cell lines derived from lung cancers but not in a large proportion of SCC. It is well known that retinoic acid has a profound effect on cell development and differentiation, particularly on the bronchial epithelium. A study using RAR-β positive and negative SCC cell lines found that the receptor positive cells grew more slowly and were less tumourigenic than their receptor negative counterparts [Howie et al, 1993]. This evidence supports the hypothesis that RAR-β may act as a TSG. C-erb-Aβ has been localised to 3p21-25 and encodes a thyroid hormone receptor. Although LOH was detected in SCLC cell lines, many did not show allele loss and a further number did not demonstrate any homozygous alterations, making c-erb-Aβ an unlikely candidate for the critical TSG [Kaye et al, 1990]. The PTP-γ gene, which maps to 3p21, is frequently deleted in both renal cell and lung cancers. Protein tyrosine phosphatases oppose the actions of protein tyrosine kinases, many of which act as oncogenes, implying that PTP genes may act as TSGs. It has been demonstrated that in 50% of primary lung cancer cell lines and tumours one allele of PTP-γ was lost [Croce, 1991, LaForgia, 1991]. Roche et al [1996] demonstrated the loss of a 3p region containing the human semaphorin IV gene, but additional data is required before the significance of this finding can be commented upon further.

Recently, attention has turned to reports of the FHIT gene (fragile histidine triad) deletion which has been found to occur in 80% of SCLC and 40% of NSCLC [Sozzi et al, 1996]. This mutation is particularly common in AC, occurring in 40% of the tumours examined. The FHIT region spans a common fragile site, FRA3B, and it is not difficult to envisage how lung cancer, a tumour closely associated with carcinogen exposure, could be prone to alterations at this site [Sozzi, 1996, Pennisi, 1996]. The FHIT gene product is similar in structure to yeast diadenosine 5', 5'''-P1, P4-tetraphosphate (Ap4A) asymmetrical hydrolase [Huang et al, 1995], and the accumulation of Ap4A may stimulate DNA synthesis and proliferation.

Many lung cancers contain LOH at 13q14 which is the retinoblastoma (Rb) locus [Friend et al, 1986]. In familial retinoblastoma, patients carry a germ line mutation of one Rb allele and the likelihood of somatic mutation in the remaining allele is high, resulting in retinal tumours in childhood. The Rb gene product is a phosphoprotein believed to be involved in the regulation of the cell cycle. Analysis of Rb DNA, RNA and protein shows an abnormality at some level in over 95% SCLC and 20% NSCLC cell lines [Birrer and Brown, 1992]. Patients with familial retinoblastoma and their heterozygous relatives have a greatly increased excess risk of lung cancer [Sandari et al, 1989]. While there is an obvious correlation between SCLC
and Rb, its role in NSCLC is not as clear cut. Studies using NSCLC primary tumours have found that up to 33% of cases contain Rb inactivation [Reissmann, 1993], but this alone is not likely to account for tumour initiation or progression. It may be that the loss of functional Rb allows the cell to leave the quiescent phase of the cell cycle and thus makes it vulnerable to the actions of mutant oncogenes or the loss of other TSGs.

Mutations of 17p13 are the commonest genetic changes observed in human malignancy. This is the locus of the p53 TSG [Finlay et al, 1989] the product of which is a phosphoprotein that complexes with the transformation proteins of various tumour viruses, such as SV40-T antigen, E1B of adenovirus and E6 of HPV 16 and 18 [Marshall, 1991]. The normal function of p53 is to inhibit cell growth and proliferation to allow damaged cells to undergo DNA repair and, if this is not possible, to initiate programmed cell death (apoptosis). If p53 function is absent, cells possessing DNA damage can continue to replicate and if the mutations are transforming and confer a growth advantage, a neoplastic clone may ensue. Mutations of p53 have been reported in over 75% of SCLC tumours and 50% of NSCLC tumours. The mutations observed in lung cancer are different to those seen in tumours from other sites and there is no obvious link with smoking history or prognosis [Mitsudomi et al, 1992]. G-T transversions like those observed in K-ras, are the most common form of mutation [Iggo, 1990, Takahashi et al, 1991]. Mutated p53 is demonstrated in approximately 40% of ACs [Harpole, 1995]. A large number show abnormal nuclear accumulation of the mutant protein and mis-sense mutations are by far the commonest [Kitamura et al, 1995]. The significance of p53 in AC development and the timing of the aberration remain unknown, as does its prognostic implication.

MCC (mutated in colonic carcinoma) and APC (adenomatous polyposis coli) are two genes located on 5q21 that have well established connections with the pathogenesis of sporadic and familial colonic carcinomas [Kinzler et al, 1991]. MCC carries somatic mutations while those in APC are in the germ line. Several studies have found deletions on 5q in human lung cancer consistent with the MCC/APC region in 20-40% NSCLC and in more than 80% of SCLC tumours [Ashton-Rickardt, 1991, D'Amico et al, 1992]. A further study isolated LOH at a locus del-27 which localises proximal to MCC/APC at 5p13-5q14 [Weiland and Bohm, 1994]. LOH at APC and del-27 was particularly apparent in SCC, showing 71% and 57% for each locus respectively. It has been shown that the APC protein is involved in the functioning of cadherin cell adhesion molecules [Rubinfield et al, 1993] and disturbances of these
molecules are associated with SCC de-differentiation and metastasis [Bohm et al, 1994]. A recent paper by White [1997] proposed a more widespread role for the APC protein in modulating cellular activity and in signal transduction pathways.

Recent findings have unmasked two possible TSG loci on chromosome 6, with LOH at both 6p and 6q occurring independently in nearly 50% of the SCLC tumours examined [Merlo, 1994]. TSGs for ovarian and breast cancers have been mapped to 6q24-25. The loss of 6p regions may include MHC loci which would give the tumour protection against immune defences and confer a selective growth advantage [Foulkes et al, 1993].

LOH on chromosome 9 occurs with similar frequency in both NSCLC and SCLC, at approximately 60%. A TSG mapping to 9p21-22 has been implicated not only in lung cancers, but in a wide variety of human malignancies including tumours of the skin, nerve cells, neuroendocrine cells and leukaemias [Merlo, 1994]. Two adjacent genes in this region were recently identified and named multiple tumour suppressor genes 1 and 2 (MTS1 & 2). MTS1 and 2 encode p16\(^{INK4}\) and p15\(^{INK4B}\) respectively. Both gene products are inhibitors of cyclin dependent kinase 4, an enzyme that acts to regulate cell division. Another interesting observation is the inhibition of the Rb gene product by p16\(^{INK4}\) and findings suggest the existence of a common p16/Rb pathway of tumour suppression [Kaye et al, 1995]. The contribution made by p15\(^{INK4B}\) is not as clear-cut. Deletions and/or mutations of MTS 1 & 2 have been detected in NSCLC cell lines and metastatic tumour samples, but not in primary tumour tissue [Packenham et al, 1995, Nakagawa et al, 1995]. The data seem to indicate the presence of a third TSG at the 9p locus [Xiao et al, 1995], the identity of which is unknown at present.

Other possible TSGs have been identified on chromosomes 8p (40% NSCLC) and 11p (45-70% NSCLC) [Emi, 1992, Bepler and Garcia-Blancer, 1994].

1.3.3 Growth factors and growth factor receptors

The overlapping functions of many growth factors are believed to be involved in the development of pulmonary AC. They include platelet derived growth factor (PDGF), insulin-like growth factor (IGF-II), basic fibroblast growth factor (bFGF) and transforming growth factor \(\alpha\) (TGF-\(\alpha\) [Tateishi et al, 1990, Takanami et al, 1996]. TGF-\(\alpha\) has been demonstrated to occur in 60% of AC cases [Tateishi et al, 1990, Carles et al, 1996]. TGF-\(\alpha\) is a ligand for
Chapter 1  Introduction to pulmonary adenocarcinoma

epidermal growth factor receptor (EGFR), the presence of which has been reported in 40-60% of AC [Tateishi et al, 1990, Johnson et al, 1995, Carles et al, 1996]. The erb-B oncogenes (erb-B1 & erb-B2/Her-2-Neu) can be activated by amplification or gene rearrangement. Erb-B1 is amplified in up to 20% of SCLC cases and protein over-expression occurs in all classes of NSCLC [Mulshine, 1993]. Amplification and over expression of erb-B2 products are also present in lung cancers, particularly AC [Guinee et al, 1995], occurring in 40-50% of this tumour cell type [Kern et al, 1990].

Pulmonary ACs expressing these genes are associated with poorer prognosis [Tateishi et al, 1990, Takanami et al, 1996]. This is not surprising as these growth factors are involved in stromal cell proliferation, angiogenesis, protease activation and mitogenic stimulation. Thus, a tumour in which such genes were expressed would be more likely to proliferate, invade and metastasise [Takanami et al, 1996].

1.3.4 Anti-metastasis and other genes

Another class of genes which may be altered in lung cancer pathogenesis includes the anti-metastasis genes. Two such genes, nm23-H1 and H2, have been described in human breast, prostate and colonic cancers [Stahl, 1991, Becker, 1993, Wang, 1993]. There is no reason to suppose that they will not also be represented in lung malignancies. Relatively few studies on nm23 in lung cancer have been published, but a recent paper by Kawakubo et al [1997] found an inverse correlation between the expression of nm23 protein and survival in AC. Genes controlling factors such as gelatinases, fibronectins and intracellular adhesion molecules, all of whom may be involved in the metastasic process, may also be abnormally expressed during tumour progression. Again, further studies are required to elucidate their role.

It can be seen that cataloguing the genetic changes associated with lung cancer pathogenesis is far from complete and those that are known are inconclusive. Every week the literature reports yet another association between lung cancer and a new genetic abnormality. There is no doubt that some of these will prove to be of great importance in the pathogenesis of lung tumours, but most will probably turn out to be non-specific, “adaptive” changes. Determining the molecular chain of events and deciding which ones are specifically involved in
tumourigenesis has profound implications for treatment, diagnosis and prevention strategies. Given the complex relationships and the potential causal factors discussed so far, it appears that the elucidation of the genetic markers that may form a multistage pathway in the development AC, and other lung cancers, will be far more difficult than was the case in colorectal cancer.

1.4. Pathology of pulmonary adenocarcinoma

As mentioned previously, the term “lung cancer” actually describes a heterogeneous group of tumours that is broadly divided into NSCLC and SCLC. The former group is by far the larger, accounting for more than 75% of all cases, and is further divided into SCC, AC and LCC sub-groups. SCLC is a separate tumour in its own right, having very distinct clinical, epidemiological and pathological features, and accounts for the remaining 25% lung cancer cases. Although these classes of tumour have defined histological characteristics, mixed or combined tumours containing features of several sub-types can occur.

1.4.1 Normal lung histology

Before any descriptions of lung pathology can given, it is first necessary to appreciate the normal histological profile of the respiratory tract. See figure 1.3. The airways are lined by pseudostratified ciliated and columnar epithelium as far as the terminal bronchioles. It must be stressed that there is no squamous epithelium in the normal tracheobronchial tree. The respiratory bronchioles and alveolar ducts possess a simple columnar epithelial lining, containing both ciliated and non-ciliated cells. The alveoli, which are invaginations of the alveolar ducts, are lined by type-1 (membranous) and type-2 (granular) epithelial cells. Within the non-ciliated epithelial population of both the respiratory bronchioles and alveoli is a special cell type known as the Clara cell. Clara cells are the progenitor cells of the bronchiolar epithelium and undergo proliferation in response to injury [Matthews and Linniola, 1988]. They also have merocrine functions and contribute to both hypophase and surfactant secretion. Clara cells are rich in p450 enzymes, making them a target for xenobiotic attack and a good candidate for neoplastic change [Serabjit-Singh et al, 1980]. Type-2 cells also possess p450s,
although not to the same degree as the Clara cells, and act as progenitor cells to replenish damaged type-1 cells.
Figure 1.3. Normal human airway histology

1.4.2 Classification of lung carcinomas

The purpose of any classification is to impose standard diagnostic criteria to ensure a high degree of consistency among histopathologists. This is vital if an accurate picture of tumour behaviour and identification of risk factors are to be established. It also has far reaching consequences for targeting research and health care provision. It has been apparent for many decades that lung cancer consisted of several discrete subtypes and was not a single tumour demonstrating extreme pleomorphism as some have claimed! This prompted the World Health Organisation (WHO) to organise a single classification system for lung cancers and the first scheme was published in 1967. Following various criticisms of the original classification criteria and also from a need to keep the system up to date, the second edition was published in 1982 and is still in use today [WHO, 1982]. It describes some 40 classes of neoplasm, 17 of which are the malignant epithelial tumours that make up “lung cancer”. See Table 1.2.

1.4.3 Diagnostic criteria for histological subtypes

1.4.3.1 Adenocarcinoma

Adenocarcinomas are found peripherally in the lung and grow within the distal airways adjacent to the chest wall. Consequently, they are much less likely to occur as surgical specimens as they are not clinically obvious until quite late in development. In the early 1960’s it was estimated that AC incidence was around 16% based on surgical material and within 10 years this figure had nearly doubled [Matthews and Linniola, 1988]. Grossly, ACs are well-circumscribed, grey-white, sub-pleural masses. Some may be fairly small and can be misinterpreted as scar tissue if surrounded by a desmoplastic (fibrous) response. Under the light microscope, ACs demonstrate the formation of glandular or papillary structures and the production of mucin. One difficulty with AC is in the distinction between primary pulmonary tumours and secondaries from other sites, such as the gut or female reproductive tract. Staining for different types of mucins can help to elucidate this problem. Another diagnostic dilemma comes from the similarity of pleural AC to malignant mesothelioma and sometimes the tumours can be indistinguishable. Although the WHO classification describes 4 types of AC, in practise it is only divided into bronchioloalveolar carcinoma (BAC) and “other”, as the remaining 3 AC types do not demonstrate any features to indicate a difference in behaviour.
Table 1.2. Revised WHO histological classification of lung tumours (1982)

I. Epithelial tumours-
   A) Benign
      1. Papillomas
         a. Squamous cell papilloma
         b. "Transitional" papilloma
      2. Adenomas
         a. Pleomorphic adenoma
         b. Monomorphic adenoma
         c. Others
   B) Dysplasia
      Carcinoma in situ
   C) Malignant
      1. Squamous cell carcinoma (epidermoid carcinoma)
         Variant:
         Spindle-cell (squamous) carcinoma
      2. Small cell carcinoma
         a. Oat cell carcinoma
         b. Intermediate cell type
         c. Combined oat cell carcinoma
      3. Adenocarcinoma
         a. Acinar adenocarcinoma
         b. Papillary adenocarcinoma
         c. Bronchiolo-alveolar carcinoma
         d. Solid carcinoma with mucin formation
      4. Large cell carcinoma
         Variants:
         a. Giant cell carcinoma
         b. Clear cell carcinoma
      5. Adenosquamous carcinoma
      6. Carcinoid tumour
      7. Bronchial gland carcinoma
         a. Adenoid cystic carcinoma
         b. Mucoepidermoid carcinoma
         c. Others
      8. Others

II. Soft tissue tumours

III. Mesothelial tumours
   A) Benign mesothelioma
   B) Malignant mesothelioma
      1. Epithelial
      2. Fibrous (spindle-cell)
      3. Biphasic

54
IV. Miscellaneous tumours
   A) Benign
   B) Malignant
      1. Carcinosarcoma
      2. Pulmonary blastoma
      3. Malignant melanoma
      4. Malignant lymphomas
      5. Others

V. Secondary tumours

VI. Unclassified tumours

VII. Tumour-like lesions
    A) Hamartoma
    B) Lymphoproliferative lesions
    C) Tumourlet
    D) Eosinophilic granuloma
    E) Sclerosing haemangioma
    F) Inflammatory pseudotumour
    G) Others
BAC is a distinct form of AC and some believe it should be treated as a lung cancer subtype in its own right. Classic BAC is a well-differentiated tumour composed of mucin-rich epithelial cells growing along pre-existing alveolar walls which may be difficult to differentiate from alveolar hyperplasia.

1.4.3.2 Squamous cell carcinoma

Squamous cell carcinoma accounts for the vast majority of surgically resected lung cancers (50-70%), but only about 35% of overall cases [Matthews and Linniola, 1988, Gatter and Dunhill, 1992]. The majority of SCCs are central in origin, that is to say that they grow from the larger, more proximal airways such as the lower trachea and main bronchi. On gross inspection these tumours appear as gray-white, bulky intraluminal masses. Microscopic examination reveals several characteristic features, including nests of pleomorphic cells, formation of keratin pearls, individual cell keratinisation and the presence of intracellular bridges. The extent of differentiation, which confers the grade of the tumour, is assessed by the degree of keratinisation.

1.4.3.3 Small cell carcinoma

Small cell lung cancers account for 25% of lung cancer cases and, like SCC, most are central tumours [Matthews and Linniola, 1988]. The majority of SCLCs occur in the upper lobes of the lung as single or multiple friable white nodules. The tumour is composed of darkly staining cells arranged in cords, nests or trabeculae. SCLCs are also called “oat cell” carcinomas by some pathologists, while others use this term to describe a specific subtype. SCLC is currently recognised as a neuro-endocrine tumour and is characterised by the presence of neurosecretory granules under electron microscopy.

1.4.3.4 Large cell carcinoma

Although large cell carcinoma is apparently a separate entity, it is often used as a “dustbin” diagnostic group for tumours that have none of the features of the groups already described and is classified primarily by exclusion. Macroscopically, these tumours tend to lie in the lung periphery as large, bulky necrotic masses. Microscopically, LCCs consist of large
cells of variable differentiation sometimes arranged in sheets. If numerous giant or clear cells are seen, the tumour may be designated as a giant or a clear cell variant. See figure 1.4.

1.4.4 Histogenesis of pulmonary adenocarcinoma

For many years research has struggled to determine how ACs develop. The main difficulty in studying ACs is due to their location in the peripheral airways. This makes it virtually impossible to observe the tumours until they become clinically apparent, by which time they are very advanced. Sampling of tissue from what may be an early or pre-malignant lesion is therefore a matter of luck. The lack of a precancerous model for AC led some people to believe the tumours arose de novo, but this is now no longer supported. Another controversial concept in AC histogenesis is the "scar cancer" theory. Most peripheral ACs contain fibrotic foci, commonly associated with pleural indentation. It was claimed that ACs developed within these sites of existing pulmonary scarring and hence were given the name "scar cancers" [Ochs et al, 1982]. Subsequent work has demonstrated that early-stage ACs do not possess these fibrous areas and it appears that the fibrosis develops after tumour formation [Noguchi et al, 1995]. It is likely that AC development is a multistep process like the adenoma-carcinoma sequence observed in colorectal tumours [Miller et al, 1988]. Adenomatous lesions may develop areas of atypical adenomatous hyperplasia (AAH), which may then become frankly malignant [Miller et al, 1988]. Many groups have used histological and cytogenetic examination as well as proliferation activity markers to monitor changes in the behaviour of these lesions. Nakayama et al [1990] reported AAH showed clonal proliferation and was either a well-differentiated AC or an adenoma. Similar results have been demonstrated by others [Shimosato et al, 1993, Rao and Fraire, 1995, Noguchi et al, 1995a, Noguchi and Shimosato, 1995b].

Another unanswered question in AC pathology is what is the cell of origin of the tumour? As discussed already, AC is divided into 4 histological subtypes but it has also been divided into 5 cytological subtypes [Shimosato et al, 1982]. Two of these cytological subtypes, Clara cell type and type-2 pneumocyte type, appear to be closely related to AAH [Shimosato et al, 1993, Nogushi et al, 1995a, Mori et al, 1996]. The majority of peripheral non-mucinous ACs share the features of these cells. The exact nature of these lesions has been
Figure 1. 4 Photomicrographs of the main types of human lung cancer.

a) Adenocarcinoma
   i] Acinar subtype

   These circular structures are glandular acini.

   ii] Papillary subtype

   Projecting masses of cells (papillary structures) are obvious in this subtype.
b) Squamous cell carcinoma

Keratin pearl formation is apparent here.

c) Small cell lung cancer

Note the nests of small, darkly staining cells.

d) Large cell cancer

Sheets of pleomorphic cells showing variable differentiation.
investigated using markers specific for Clara cells and type-2 pneumocytes. No differences were established between AAH and either type of AC using antibodies to urine protein 1 (Clara cell marker), SP-A (Type-2 cell marker), carcinoembryonic antigen (AC marker) and p450s (markers for both cell types), or from morphometric analysis [Mori et al., 1996]. This has been interpreted as supporting the assumption that AAH is a neoplastic lesion closely related to AC or that it is a common precursor of these two types of AC. These lesions often co-exist and show similar ultrastructural features with electron microscopy [Jaques and Currie, 1977].
Chapter 1  Introduction to pulmonary adenocarcinoma

1.5. Clinical features of pulmonary adenocarcinoma and other types of lung cancer

1.5.1 Clinical symptoms

Lung cancer may present in a variety of ways depending on the tumour type, size and location. However, most of the symptoms are present to varying degrees in all tumour cell types. The most common symptoms include persistent cough, chest pain, haemoptysis (coughing up blood) and dyspnoea (breathlessness). Lung cancers frequently ulcerate the bronchus causing varying degrees of haemoptysis in up to 50% of patients. Tumours growing within the lumen of an airway will obviously cause an obstruction, this results in dyspnoea and in the collapse of the lung distally. Lung collapse results in the retention of secretions and secondary infection which may then cause pneumonia or a lung abscess, depending on the bacterial agent involved. Adenocarcinomas also cause a particularly distressing symptom called mucorrhoea. This is the sudden release of large amounts of mucus from the tumour and results in intense dyspnoea. As lung tumours spread, local invasion and metastasis produce further symptoms. The involvement of the tumour in various neural structures in the thorax can produce hoarseness and impaired diaphragmatic function. Compression of the oesophagus leads to dysphagia (swallowing difficulty), compression of the inferior vena cava (IVC) may also occur. Lymphatic spread gives rise to fluid collection in the pleural space, called a pleural effusion, which impairs lung function. The spread of the tumour to distant sites via the blood stream is very common, organs affected including the liver, bone, adrenal glands and brain. Some types of lung cancer, particularly SCLC, produce ectopic hormones. These hormones are responsible for many of the para-neoplastic effects observed in this disease, such as inappropriate anti-diuretic hormone secretion which causes retention of water.

1.5.2 Diagnostic methods

The definitive diagnosis of lung cancer can only be made histologically, no matter how seemingly suggestive the symptoms or chest x-rays may be. Diagnostic samples may be obtained in several ways, from sputum collection to open surgery. Cytology specimens, from
sputum or pleural effusions, are perhaps the simplest starting point. If more than one sample is examined, up to 70% of lung cancers can be diagnosed in this way [Gatter et al., 1992]. The fine-needle aspiration biopsy is another well-established diagnostic technique and needle aspiration is used to remove a sample of the tumour under radiological control. It has the advantage over cytology of preserving more of the tumour architecture. However, the most reliable method is the examination of sections of the tumour tissue itself and this has to be achieved by surgical biopsy. Today, this is performed with the flexible bronchoscope (FB), a fibre-optic tube that is passed down the naso-pharynx and into the bronchial tree. The FB has enabled sampling of even peripheral lesions previously only accessible by open chest surgery. It is not without its drawbacks, however. Samples obtained are small and frequently do not contain diagnostic material, and many biopsies are crushed or lost. Such specimens also sample only very small areas which may lead to misclassification in the case of a mixed tumour. Open chest surgery is the most extreme method of tissue collection but is the most reliable! Occasionally, wedge biopsies containing small lesions may be removed, or a biopsy for frozen section analysis may be taken so the result can be known and either partial/complete lobectomy (removal of one lobe of the lung) or pneumonectomy (removal of the whole lung) may follow immediately.

1.5.3 Treatment and prognosis

Lung cancer is a death sentence. It is an insidious disease that carries a poor prognosis because what is “clinically early is biologically late”. By the time the patient presents to a doctor there is really very little that can be done in most cases. More than 3/4 of lung cancers are untreatable at the time of diagnosis [Shepherd, 1994]. Apart from the histological grading of tumours according to their degree of differentiation, they are also assessed clinically by the degree of spread, or stage. Lung tumour staging is based on the Tumour/ Nodes/ Metastases (TNM) clinical classification system; “T” refers to tumour size, “N” indicates lymphatic spread and “M” denotes the presence/ absence of secondary tumours. Clinical staging involves a complete history and physical examination as well as laboratory investigations, thoracic imaging techniques and invasive procedures, such as mediastinoscopy, lymph node biopsy and thoracotomy. See table 1.3. Approximately 80% of lung cancer cases are at stage IIIIB or IV.
at diagnosis, hence the reason for the rather gloomy opening statement. Only stage I and II
tumours are suitable for treatment, this equals 6-15% of NSCLC and <3% SCLC cases.

Table 1.3. Staging of lung cancer

a. TNM Clinical Classification

T- Primary tumour
TX Cannot be assessed, or tumour proven by presence in cytology specimen only
TO No evidence of primary tumour
Tis Carcinoma in situ
T1 <3 cm in greatest dimension
T2 >3 cm in greatest dimension/ involves main bronchus/ involves visceral pleura/
distal lung collapse
T3 Tumour any size & locally invasive (chest wall, pleura, pericardium)
T4 Tumour any size & spread to heart, great vessels, oesophagus, spine, trachea ... or
tumour with malignant pleural effusion

N- Regional lymph nodes
NX Cannot be assessed
NO No lymph node metastasis
N1 Metastasis in ipsilateral local lymph nodes
N2 Metastasis in ipsilateral mediastinal/ subcarinal lymph nodes
N3 Metastasis in contralateral lymph nodes

M- Metastases
MO No metastases
M1 Metastases present

Adapted from: UICC International Union Against Cancer. Hermanek P & Sobin L.N. (eds)
### Chapter 1  Introduction to pulmonary adenocarcinoma

#### b. Stage grouping

<table>
<thead>
<tr>
<th>Occult carcinoma</th>
<th>TX</th>
<th>NO</th>
<th>MO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage O</td>
<td>Tis</td>
<td>NO</td>
<td>MO</td>
</tr>
<tr>
<td>Stage I</td>
<td>T1</td>
<td>NO</td>
<td>MO</td>
</tr>
<tr>
<td></td>
<td>T2</td>
<td>NO</td>
<td>MO</td>
</tr>
<tr>
<td>Stage II</td>
<td>T1</td>
<td>N1</td>
<td>MO</td>
</tr>
<tr>
<td></td>
<td>T2</td>
<td>N1</td>
<td>MO</td>
</tr>
<tr>
<td>Stage IIIA</td>
<td>T1</td>
<td>N2</td>
<td>MO</td>
</tr>
<tr>
<td></td>
<td>T2</td>
<td>N2</td>
<td>MO</td>
</tr>
<tr>
<td></td>
<td>T3</td>
<td>NO,1,2</td>
<td>MO</td>
</tr>
<tr>
<td>Stage IIIB*</td>
<td>Any T</td>
<td>N3</td>
<td>MO</td>
</tr>
<tr>
<td></td>
<td>T4</td>
<td>Any N</td>
<td>MO</td>
</tr>
<tr>
<td>Stage IV*</td>
<td>Any T</td>
<td>Any N</td>
<td>M1</td>
</tr>
</tbody>
</table>

[* = Un救治able]
Chapter 1 Introduction to pulmonary adenocarcinoma

Lung cancer treatment varies according to cell type, stage and the general condition of the patient. The presence of concomitant pulmonary or cardiovascular disease precludes surgical intervention as the patient probably would not survive the procedure. Treatment options for lung cancer patients are limited to combinations of the following: surgical resection, radiation therapy (DXT), chemotherapy, immunotherapy and laser therapy. Surgical resection is limited to early stage (I & II) NSCLC without lymph node or metastatic involvement, although some studies have included stage III and IV patients [Hsu, 1996, Shepherd, 1994]. The 5 year survival rate in the former group can be as high as 70%. The use of surgery for SCLC is, in general, uncommon due to the frequent occurrence of micro/macrometastatic tumours at presentation. DXT can be applied to all forms of lung cancer, although different cell types differ in their radiosensitivity; SCC and SCLC often show good responses to moderate doses of 3500-5500 rad, while AC and LCC respond less well [Shepherd, 1994]. Palliative DXT is also given to patients with more advanced disease to relieve some of the symptoms caused by tumour bulk, such as bronchial obstruction and IVC compression. Chemotherapy is the main method of SCLC treatment and is more effective when drugs are given in combination. The most effective regimens include cyclophosphamide/doxorubicin/cisplatin (CAP) and cyclophosphamide/doxorubicin/methotrexate/procarbazine (CAMP) [Shepherd, 1994, Limsila, et al 1996]. Combination chemotherapy is now also being used to treat NSCLC. The taxoids paclitaxel and docetaxel are two new agents which have shown promising single-agent activity in non-small cell cancer [Rigas, 1997]. Immunotherapy is still in its infancy but has the potential to yield some exciting possibilities in the future. It currently includes treatment with non-specific antigens, such as bacillus Calmette-Guerin (BCG), to stimulate the immune system, as well as direct treatment with interleukins, interferons, colony stimulating factors and monoclonal antibodies raised against lung tumour antigens [Bunn, 1994, Shepherd, 1994, An et al, 1996]. Photodynamic laser therapy is proving very effective in the treatment of early tumours and carcinoma in situ and the use of fluorescent laser bronchoscopy also allows visualisation of dysplastic lesions not visible with conventional white light imaging [Furuse et al, 1993]. The vastly superior optics of video-assisted thoracic surgery has rendered older thoracoscopy equipment obsolete. This system has allowed greater accuracy in biopsy sampling and is proving particularly useful in the excision of solitary tumour nodules [Goldstraw, 1997].
Prognosis is poor regardless of grade, stage or tumour type. See table 1.4. The average life expectancy is 3-6 months from diagnosis and 85-95% of all patients will die from the disease within 5 years. These data make it imperative that something is done to improve the situation and high-light the problem of treating something that cannot be seen until it is too late. It has been estimated that within an average-sized health district of 250,000 people, there will be at least 75 new cases of lung cancer each year. Assuming an average life span of 75 years, this will cost 1700 years in lost life and approximately £1 million in palliation/treatment. The epidemic of lung cancer that is sweeping the globe is carrying with it a high cost both in terms of money and, more importantly, in terms of human suffering.

Unlike the advances made in the diagnosis and treatment of other malignancies, such as childhood cancers, leukaemias and lymphomas, and for some solid tumours, such as colorectal carcinoma, the prognosis for lung cancer sufferers has remained largely unchanged for over two decades. What can be done to combat this problem? How can current therapies be improved? What role will molecular biology play in the future of lung cancer research?

Table 1.4. Prognosis of lung cancer subtypes

<table>
<thead>
<tr>
<th>Tumour type</th>
<th>Treatment</th>
<th>% cases</th>
<th>Survival</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSLC</td>
<td>Resection</td>
<td>6-15%</td>
<td>50% at 2 yrs if localised</td>
</tr>
<tr>
<td></td>
<td>DXT</td>
<td>5-15%</td>
<td>10% at 2 yrs if spread</td>
</tr>
<tr>
<td></td>
<td>Palliation</td>
<td>&gt;70%</td>
<td></td>
</tr>
<tr>
<td>SCLC</td>
<td>Resection</td>
<td>&lt;3%</td>
<td>Median 12 months</td>
</tr>
<tr>
<td></td>
<td>DXT</td>
<td>30%</td>
<td>for all treated cases &amp;</td>
</tr>
<tr>
<td></td>
<td>Chemo</td>
<td>25-50%</td>
<td>1-3 months if untreated</td>
</tr>
</tbody>
</table>
1.6. The mouse model of pulmonary adenocarcinoma

One of the problems of studying human pulmonary adenocarcinoma is the constant failure to identify the premalignant and early tumour stages. Consequently, knowledge of the genetic changes that occur during these important phases of tumourigenesis remains largely speculative. The development of the mouse model of pulmonary adenocarcinoma (AC) has been invaluable in helping scientists learn more about these very aspects and applying the data to the human disease.

1.6.1 History of the development of the mouse model

The first description of spontaneous lung tumours occurring in mice was published by Livingood in 1896. The first use of inbred mouse strains to examine inherited lung tumour susceptibility was demonstrated by Clara Lynch 30 years later [Lynch, 1926]. Lynch noticed that the frequency of spontaneous lung tumours in wild mice was less than 3%, but the inbred strains showed varying frequencies. The inbred strain that is most commonly used to study pulmonary AC is the strain A mouse, established in 1921 [Dragani et al, 1995]. It was observed that a high percentage of these animals developed spontaneous lung tumours [Strong, 1936], and virtually 100% developed the same tumours following dosing with a carcinogen [Andervont, 1937]. The mouse model has been applied to several areas of study, the main use being as a bioassay for certain classes of chemical carcinogens [Shimkin and Stoner, 1975]. It has also been used to examine tumour susceptibility determinants [Festing et al, 1994] and the modulation of tumour development by both tumour promoters [Witschi et al, 1977] and chemopreventative agents [Wattenberg, 1973].

1.6.2 Similarity to human pulmonary adenocarcinoma

Murine pulmonary AC shows sufficient similarities to the human form of the tumour to warrant its use as a substitute for studying premalignant and early stage disease. These similarities can be classed into 3 areas; histology, susceptibility factors and changes in gene expression.
1.6.2.1 Histology

The histological appearance of murine lung tumours are similar to the AC observed in the human lung, most closely resembling the bronchioloalveolar carcinoma subtype (BAC) [Malkinson, 1992, Dragani et al, 1995]. Spontaneous and induced mouse lung tumours have similar histological features and can be divided into two subtypes, alveolar (or solid) and papillary. The alveolar/solid subtype consists of oval/cuboidal cells forming cords or nests around the alveoli [Dragani et al, 1995]. They have ultrastructural features similar to those recorded in type-2 cells [Grady and Stewart, 1940] and also produce surfactant [Snyder et al, 1973, Malkinson, 1989]. The papillary tumours have ultrastructural features resembling the Clara cell [Kaufmann et al, 1979] and produce enzymes associated with Clara cell function [Gunning et al, 1991]. Papillary tumours tend to develop more slowly than the alveolar tumours and are more likely to be malignant [Thaete et al, 1987]. The cell of origin of the mouse tumours has long been debated and the matter is still contentious, with support being divided approximately equally between the type 2 cell and the Clara cell [Malkinson, 1992, Dragani et al, 1995].

1.6.2.2 Susceptibility factors

The genetic basis for the susceptibility of human AC has already been discussed in chapter 1. Malkinson et al [1985] identified three groups of genes that control pulmonary adenoma susceptibility in the mouse, known as the PAS factors. These include the H-2 histocompatibility complex [Faraldo et al, 1979] and polymorphisms in the K-ras proto-oncogene [Ryan et al, 1987]. In addition, other host factors such as age and immune status also affect murine pulmonary adenoma development [Beer and Malkinson, 1985].

1.6.2.3 Changes in gene expression

Mutations in K-ras occur in human pulmonary AC at a frequency of between 30-56% [Rodenhuis, 1992, Mills et al, 1995]. High rates of K-ras mutation (77-95%) are observed in spontaneous lung tumours from susceptible mouse strains [You et al, 1989, Li et al, 1994a]. Mutation rates occurring in spontaneous tumours from resistant strains are much lower, ranging from 10-43% [Candrian et al, 1991, Devereux et al, 1991]. Chemically induced lung tumours from susceptible and resistant strains show similar rates of K-ras mutation, from 80-100%, indicating that susceptibility does not correlate with mutability of K-ras [Devereux et
Chapter 1  Introduction to pulmonary adenocarcinoma

1.6.1 K-ras mutations

The K-ras mutations target codons 12, 13 and 61, just as in humans. The 5' polymorphism that occurs in K-ras from susceptible strains may influence the ability of the gene to undergo subsequent mutation [Malkinson et al., 1994, Chen et al. 1994]. It is believed that K-ras mutation is an early event in mouse lung adenoma development, as it occurs in 85% of hyperplastic lesions of strain A mice treated with NNK [Belinsky et al., 1992].

Mutations in p53 are also fairly common in human pulmonary AC [Harpole et al., 1995]. The presence of mutations of p53 (exons 5-8) in chemically induced lung tumours from several strains of mice has been investigated by several groups [Chen et al., 1993, Li et al., 1994a]. However, p53 mutations could not be demonstrated in any of the studies.

Loss of heterozygosity (LOH) has been observed for mouse chromosomes 4, 11 and 14, which correspond to human chromosomes 9p, 17p (p53) and 13q (Rb), respectively [Dragani et al., 1995]. The chromosome 4 LOH localised near to the α-interferon locus and occurred in 48% of pulmonary AC studied [Dragani et al., 1995]. In humans, LOH at 9p occurs with a frequency of 36-63% [Merlo, 1994] and tumour suppressor genes MTS 1 and 2 map to this region. LOH of chromosome 11 surrounding the p53 gene have been demonstrated in 13-20% of certain chemically induced lung tumours from O mice [Hegi et al., 1993]. LOH in human AC has been reported to occur at frequencies from 28-70% [Dragani et al., 1995]. On chromosome 14, LOH of the Rb gene is infrequent in mouse lung tumours [Hegi et al., 1993], whereas in humans it occurs more commonly, particularly in SCLC [Birrer and Brown, 1992].

1.6.3 Uses of the mouse model

Several experimental approaches are available to study the mouse lung tumour system. The first is the ability to obtain premalignant tissue at any time, following the induction of lung tumour development with a suitable chemical carcinogen. Following dosing, hyperplastic nodules can be seen as early as +2 weeks, small benign lesions are present at +10 weeks, with adenocarcinomas developing at approximately +8 months [Malkinson, 1992].

The use of recombinant inbred (RI) strains identifies strain distribution patterns which examine single gene effects. RI strains result from mating different inbred strains (A x B) to give an F1 hybrid generation. These F1 offspring are then mated via brother x sister crosses for
at least 20 generations, producing an RI strain with homozygous alleles representing one or other of the original parental strains (A or B). If phenotypes are observed that vary from those found in A or B, it implies that the trait is controlled by more than a single gene [Nesbitt, 1984]. RI strains from susceptible strain A and resistant strain C57BL were used to identify the PAS factors [Malkinson et al, 1985].

A recently developed application of the mouse model has been the use of transgenic animals. Constructs containing Clara cell protein promoter/ SV40- T antigen [Sandmoller et al, 1995], surfactant apoprotein promoter/TGF-α [Glasser et al, 1994] and albumin promoter/H-ras [Sandgren et al, 1989] have all been used to produce transgenic animals. This has allowed the acceleration of tumour development, with malignancies now occurring in less than 8 weeks instead of >8 months [Maronpot et al, 1991]. Transgenic lines now offer a new approach to the study of potential chemotherapeutic and chemopreventive agents [Dragani et al, 1995].

1.6.4 Disadvantages of the mouse model

Despite the obvious advantages this system allows in furthering the study of human pulmonary AC, several caveats much be remembered. Firstly, mice are not human! In general terms, the biochemical, physiological and immunological processes occurring in the mouse will not be the same as those in the human. Thus, the cellular environment will differ for each species. However, the malignant lung tumours that mice develop do resemble some types of human pulmonary AC, but they are not identical. The Clara cell frequency and distribution is different in the mouse than in a human. These cells account for between 50-60% of the cells in the mouse airway, a much higher figure than in the human [Pack et al, 1981]. In humans, Clara cells are confined to the distal airways but in the mouse they also predominate in the upper airways [Pack et al, 1980]. Finally, the behaviour of mouse lung tumours is not as aggressive as their human counterparts. Murine pulmonary tumours grow by expansion, coalescence and infiltration, but the additional tumour burden frequently reaches a lethal level before the lesions have progressed to full malignancy, let alone formed metastases [Foley et al, 1991].
1.7. The future of lung cancer research

1.7.1 Screening strategies

Lung cancer screening is currently heading in two directions; firstly, towards the identification of susceptible individuals who have not yet developed the disease and secondly, towards the identification of those who already have the disease at its earliest stages. Oncogene mutations have been detected in the urine of patients with bladder tumours [Sidransky et al., 1992] and in the faecal material from patients with colonic cancers [Sidransky et al., 1991], so why not use sputum samples from lung cancer patients? This is exactly what has been attempted by several groups over the past five years. A retrospective study of K-ras and p53 mutations in sputum was reported by Mao et al. [1994] using a PCR-ASO hybridisation based assay. The group found that 10/15 patients had primary lung cancers containing one of the mutations, of which 8/10 showed identical findings in sputum samples. The earliest detection occurred in a sample taken 12 months prior to clinical diagnosis. Similar findings were obtained when cytological specimens were tested for the presence of K-ras mutation [Kiaris et al., 1995]. The problem with these studies stems from the relative paucity of malignant cells in comparison to normal “background” cells and the fact that only a fraction of malignant cells are carrying the ras mutation. This results in competition during amplification which diminishes any signal from the malignant target population. This problem was overcome with a new and highly sensitive PCR-based technique called PCR-PIREMA, discussed earlier, which almost doubled the frequency of mutation finding to 56% [Mills et al., 1995]. If this technique could be applied to a marker that was found in the majority of lung cancers of a given cell type it would form the basis of a very powerful screening assay.

1.7.2 New therapies

Several new therapy options are being examined at the moment. The first gene therapy studies were reported by Roth's group using K-ras antisense constructs in animal subjects [Mukhopadhyay et al., 1991]. By 1996, the same technology was being applied to humans, this time by the transfer of wild-type p53 via bronchoscopic injection into the tumour [Roth et al., 1996]. Of the nine patients selected for the study, 3 showed signs of tumour regression and a
Chapter 1  Introduction to pulmonary adenocarcinoma

Further 3 tumours demonstrated growth stabilisation following treatment. The findings were limited to local tumour within endobronchial lesions only. The use of wild-type p53 has also been reported as an adjuvant to DXT/ chemotherapy in the induction of apoptosis in tumours [Fujiwara et al, 1994] and may form the basis of new combination therapies to control local tumour growth. A recent study by Heike et al [1997] reported an attempt at gene therapy in the mouse using an adenoviral vector carrying an interleukin 2 gene construct. It was hoped that this construct would induce anti-tumour activity and decrease tumourigenicity. While the treatment was effective on local tumours, it had no effect on distant disease.

Slightly more conventional trials are being undertaken with chemopreventive agents. Retinoids, such as isotretinoin and retinol palmitate, are being used to prevent recurrence of malignancy following curative resection [Pastorino and Infante, 1993, Hong et al, 1990]. Vitamin A analogues have also been used in the prevention of squamous metaplasia [Geradts et al, 1993], although preliminary results have been mixed. Recently, novel retinoid-related compounds have been found that induce apoptosis in animal lung cancer cells in vivo [Lu, 1997]. Naturally occurring isothiocyanates, such as phenethyl isothiocyanate and benzyl isothiocyanate, have been found to inhibit lung tumour induction in rodents by the constituents of tobacco smoke [Hecht, 1995]. The antimitogenic and anticarcinogenic properties of N-acetylcysteine (NAC), a synthetic precursor of cellular of intracellular cysteine and glutathione, have long been known. The preliminary data from the EUROSCAN phase III trials, which compared the relative efficacy of retinol palmitate and NAC in ~2,500 patients with previous respiratory tract cancers, are currently being analysed [De Vries et al, 1991]. Other groups are working on chemotherapeutic agents, such as topoisomerase inhibitors, cytokines, interleukins and growth factors [Bunn, 1994, Rusch and Dmitrovsky, 1995].

Hopefully, the volume of lung cancer research will reach a critical mass in the near future and will begin to provide answers to some of the questions posed. Perhaps, we will even see the beginning of the end of this modern day plague within our professional lifetimes.
1.8. Questions addressed by this study

1.8.1 Why study pulmonary adenocarcinoma?

It is apparent from what has been discussed already that far more is unknown about the development of pulmonary AC, and other lung cancers, than is known! Although many breakthroughs have occurred in our understanding of the mechanisms of this disease, we still have a very long way to go before this can be translated into any practical aid for the many thousands of people who suffer from these tumours. Unlike other forms of cancer, where there have been significant improvements in treatment and prognosis over the past 20 years, such as leukaemias, lymphomas, gastro-intestinal cancer and tumours of the reproductive system, lung cancer prognosis has not improved and, if anything, its mortality has risen. Globally, lung cancer has now reached epidemic proportions over the past 50 years and the incidence is set to rise well into the next century. If the incidence patterns continue as they are now, the majority of lung cancer cases will be of the AC subtype. We know that the vast majority of lung cancers are related to exposure to tobacco smoke. We do not know why only 15-20% of smokers develop the disease or why so many non-smokers develop lung cancer, particularly AC. We know that the proportion of cases of AC is rising, but we do not know why. We know that more women are getting this form of lung cancer, but we do not know why. We know that the key to understanding pulmonary AC, and other lung cancers, lies within the molecular mechanisms of carcinogenesis. We do not know what these mechanisms are. We know there must be multiple genetic steps involved in the development of all lung cancers, including AC. We have identified numerous candidate genetic markers, none of which are significant in all tumours.

Thus, the elucidation of the genetic abnormalities involved in the pathogenesis of pulmonary AC, and other lung cancer subtypes, is of great importance. It would allow the development of screening tests to define individuals at risk of developing AC and other lung tumours. It would provide diagnostic aids to augment or replace those currently in use, and supply new indicators for selecting therapies such as chemoprevention trials, adjuvant therapy and advanced disease therapy. In time, it may even form the basis of new treatments, such as gene replacement.

The question is how?
1.8.2 How to study pulmonary adenocarcinoma?

How do we determine the steps in a complex pathway of events that may not only be different for each tumour type, but that may demonstrate variation according to sex, carcinogen exposure, and a host of other factors specific to each individual?

Identifying genes involved in human disease is a complex and laborious process, often taking many years of study. One way is to identify several candidate genes, implicated in the disease process. Techniques to accomplish this involve positional and functional cloning [Strachan and Read, 1996]. In positional cloning nothing is known about a gene except its sub-chromosomal location. This is achieved by linkage analysis with disease markers, loss of heterozygosity screening to identify chromosomal deletions associated with the disease, and cytogenetic analysis to determine other chromosomal abnormalities present in disease samples. Functional cloning uses information about the function of an unknown disease gene as a way of identifying it. It is sometimes possible to obtain partially pure gene products, particularly for genes involved in hereditary metabolic disorders. If sufficient protein can be recovered the amino acid sequence can be determined and used to produce partially degenerate oligonucleotides. These are then used to screen cDNA and genomic DNA libraries to isolate the full length clone. Antibodies raised against the protein can also be used to screen expression cDNA libraries. As already stated, these procedures can take many years, if not decades, before a single gene is isolated.

The other way of disease-gene identification involves “going in blind”, that is to say, not having any idea of the identity or function of the gene. Methods using this approach are based on the comparison of mRNA populations from normal and disease samples, and include subtractive and differential hybridisation [Wang and Brown, 1991, Sargent, 1987]. These techniques identify mRNA species unique to a given sample and tend to be a lot quicker than the cloning approaches described. Their main disadvantage is the requirement of large amounts of mRNA and a relatively high incidence of false positives. Recently, a new PCR-based technique has been developed which also compares mRNA populations using only small amounts of starting material. It is called differential display reverse transcription PCR (DDRT-PCR) and forms the main experimental method of this study.
The working hypothesis of this study was that by using DDRT-PCR to compare the mRNA populations of matched normal lung and pulmonary adenocarcinoma samples, it may be possible to discover differences in the expression of significant genes between these samples. Some of these differences may be due to the expression of genes that are directly involved in the process of tumourigenesis.

1.8.3 Aims of this study

The rationale behind this study was to attempt to learn more about the genes expressed in human pulmonary adenocarcinoma using differential display reverse transcription PCR. By comparing mRNA populations from normal lung and pulmonary adenocarcinoma, the identities of novel genes that may be involved in tumourigenesis may be determined. The experimental method was standardised for use with lung RNA, initially with human cell lines and then with human tissues. Once candidate genes have been identified, the cell populations in which these genes are expressed could be identified. The range of expression could then be determined in non-pulmonary tissues, from both human and animal sources. These techniques could also be applied to pre-neoplastic and metastatic tissues to examine expression at different stages of the neoplastic pathway.
Chapter 2 -

Introduction to differential display
reverse transcription PCR
2.1 Principles and development of DDRT-PCR

The comparison of mRNA expression between normal and abnormal cells provides valuable information concerning the activity of genes within these cells. Some of the alterations in gene expression may result in the development of pathological processes and of particular interest are the changes involved in carcinogenesis. Differential and subtractive hybridisation methods were the principal techniques used to examine mRNA transcripts [Wang and Brown, 1991, Sargent, 1987]. Differential hybridisation, also known as plus/minus screening, allows a gene-by-gene comparison of the mRNAs between two samples by making a cDNA library from one sample and probing copies of it with labelled cDNA probes made from both samples. Subtractive hybridisation is similar, except that cDNA probes made from one sample (test) are hybridised against an excess of mRNA from the other sample (driver) to remove any common cDNA-mRNA sequences, leaving those which are unique to remain single-stranded. The cDNA-mRNA hybrids are removed and the single-stranded, or subtracted cDNAs, are recovered and cloned. These methods have several drawbacks. They are time-consuming and labour-intensive, with 2 months being required to isolate clones from cells by subtractive hybridisation. They are difficult to establish and only when the final results are known is it apparent whether the procedure has succeeded or not. Large amounts of RNA are required (>200 µg), rare mRNA transcripts may be undetected and when results are obtained they may not be reproducible. An alternative method was required which could detect all mRNA species expressed in a cell and which identified increases and decreases in gene regulation activity.

In 1990, a technique using DNA fingerprinting with arbitrary primers was developed and named RAP-PCR (random arbitrary priming) [Welsh and McClelland, 1990, Williams et al, 1990]. The method was used to identify DNA fingerprints in colorectal carcinoma samples by Peinado et al [1992]. Later the same year an extension of the DNA protocol was developed for use with RNA [Welsh et al, 1992]. At the same time Liang and Pardee published a paper describing their independent development of a related, yet subtly different, protocol for the differential display of mRNA using PCR and the method that was to become known as differential display reverse transcription -PCR was created (DDRT-PCR) [Liang and Pardee, 1992]. For some reason it was DDRT-PCR, rather than RNA RAP-PCR, that captured the imagination of the research population and since its conception the literature on DDRT-PCR
has increased exponentially. The originators of the technique made the assumption that only 10-15% of the estimated 100,000 genes present within each cell are ever expressed at any one time and that any one of the 10,000-15,000 resulting mRNA transcripts can be reverse transcribed and amplified by PCR. The underlying principle of the technique is to divide the mRNAs into subgroups and produce partial cDNA sequences by reverse transcription which are then amplified using PCR. Radiolabelled dNTP is incorporated into the PCR products enabling direct visualisation of the banding pattern on a sequencing gel. Comparison of the patterns obtained from control and test RNA samples allows the identification of possible differentially expressed cDNAs. See figure 2.1. DNA from any band of interest can be extracted from the gel and cloned for use as a probe for northern and Southern blotting, which confirms differential expression. Differentially expressed clones can then be sequenced and used to provide full length transcripts from cDNA libraries or by using RACE-PCR, see figure 2.2.

The process begins with the isolation of total RNA from either cell cultures or tissue samples. The exact method used for RNA extraction is not important providing it yields good quality, undegraded total RNA. Poly(A) RNA can also be used for DDRT-PCR but the use of oligo(dT) primers frequently gives a background smear caused by priming without an anchor primer to the mRNA poly(A) tails (see below). However the RNA is made, one vital procedure is its treatment with DNase I to remove any contaminating chromosomal DNA and thereby minimise false-positives and non-specific banding (see below). The integrity of the RNA should be checked by agarose gel electrophoresis before use in DDRT-PCR. The technique is so sensitive that as little as 200 ng RNA per RT reaction are required.

The 3' primer is an oligo(dT) sequence which binds to the poly(A) tail of the mRNA and is known as the anchor primer. This primer also contains 2 additional bases at its 3' end which provide specificity, i.e. \( T_x M_N \) [Liang and Pardee, 1992, Liang et al 1993, Bauer, 1993]. Omitting \( T \) as the penultimate base, there are 12 possible combinations of \( M_N \), therefore each primer will anneal to approximately \( \frac{1}{12} \)th of the total mRNAs present. The anchor primer oligo(dT) usually contains 11-12 \( T \) bases. Increasing or decreasing this number gives either minimal banding or smearing on the gel. This emphasises the importance of the 3' bases (MN) in the anchoring specificity at the beginning of the poly(A) tail. The penultimate base on the anchor primer (M) can exhibit marked degeneracy during the RT priming step, whereas the last base (N) dominates specificity. This means that 4 degenerate \( T_x M_N \) primers differing only in
Figure 2.1. Principles of differential display RT-PCR

I. REVERSE TRANSCRIPTION

5'\text{T}_{12}\text{MN}-3' \\
dNTPs \\
M-MLV ENZYME

II. PCR AMPLIFICATION

5'\text{AGCCAGCGAA}-3' (AP-1 PRIMER) \\
5'\text{T}_{12}\text{MN}-3' \\
dNTPs/^{35}\text{S}/AMPLITAQ ENZYME

III. DENATURED PAGE

RNA A | B
| | | | | | | |

DIFFERENTIALLY EXPRESSED BAND

i. Individual mRNAs are reverse transcribed using each anchor primer
ii. Resulting cDNAs are amplified by arbitrary primers during PCR
iii. PCR products are separated and visualised on a sequencing gel

KEY: 
mRNAs

cDNAs

PCR products

(Redrawn from Biogene handbook)
1. Remove bands from DDRT gel, extract cDNA and reamplify

2. Clone PCR products

3. Verify differential expression by northern hybridisation

4. Screen against cDNA library to obtain full length cDNA and sequence

5. Use probe for \textit{in situ} hybridisation and determine range of expression
their last 3' base can be used, which reduces the number of RT reactions from 12 to 4 per RNA sample [Liang et al, 1993].

The 5' primers, known as arbitrary primers (AP), have arbitrary sequences and will anneal at random distances from the poly(A) tails thus producing bands of varying lengths when run out on a polyacrylamide gel. The APs are selected to produce DNA sequences of <500 bp as this length can be visualised and resolved on a sequencing gel. Selection of the AP group proved to be more complicated than for the anchor primers. It was found that APs of <9 bases gave poor results and an optimum length of 10 bases was eventually determined. These 10-mers, in conjunction with the anchor primers, amplified more bands than statistically predicted and behaved more like 6- or 7-mers [Liang and Pardee, 1992, Bauer, 1993]. This phenomenon was explained by degenerate hybridisation during initial PCR and an increase in sub-sequences capable of stable hybridisation. It was calculated that 20 arbitrary 10-mers priming as 6-/ 7-mers should cover all mRNAs upstream of the anchor primers [Liang and Pardee, 1992]. In selecting AP sequences factors such as GC:AT values and degree of self-complementarity must be considered. The AP sequences from early DDRT-PCR literature have now become standard and are used by most researchers. During the initial phase of the reaction the RNA target is reverse transcribed separately with each anchor primer. The anchor primers are then re-used in the second phase of the reaction when the resulting cDNAs are amplified with each degenerate AP 10-mer in turn. Each primer combination yields an average of 70-150 bands per lane of sequencing gel and with 40 lanes available per gel, theoretically the entire mRNA population could be displayed on 4 such gels. Banding patterns were found to be 95% reproducible when the same primer combination and RNA samples were used in more than 3 independent experiments [Liang and Pardee, 1992].

In addition to selection of primer sequences, several other reaction conditions have a profound effect on the outcome of DDRT-PCR; PCR annealing temperature, dNTP concentration and Mg²⁺ concentration [Liang and Pardee,1992, Liang et al, 1993, Bauer, 1993]. Experiments evaluating each of these variables determined annealing temperature as the most influential factor. Values <40° C gave bands at virtually all positions of the gel, while raising the temperature to >42° C caused the number of bands to fall sharply. Similar results were demonstrated for dNTP and Mg²⁺ concentrations, with optimal values determined at 2 μM and 1.25 μM respectively.
2.2 Problems and modifications of DDRT-PCR

Initial reaction to the arrival of this technique was good. At last here was a method to allow investigation of gene expression in a way not previously thought possible. Minimal amounts of starting material were required, total reaction time from PCR to probing blots was as little as 14 days, the process could be checked at each step rather than only at the end, it was reproducible, sensitive and avoided problems of underrepresenting rare mRNAs. DDRT-PCR seemed almost too good to be true! However, as larger numbers of researchers applied the technique to their studies, it became apparent that DDRT-PCR was not quite as infallible as had been previously thought. The main problems are the number of false positive bands, non-specific background "noise" and the fact that the cDNA tags obtained are usually short (<300 bp) and were often situated in the 3¨untranslated region of the mRNA.

The number of false positive bands obtained from DDRT-PCR was originally thought to be approximately 5%, but is now estimated to be in the region of 50-70% [Debouck, 1995, Sunday, 1995]. This high rate has been attributed to the presence of multiple cDNA fragments present in excised bands with estimates of at least 3 co-segregating cDNAs at each band position. These cDNAs may be caused by several factors; independent cDNA fragments of the same molecular weight, separated strands of the same unique double-stranded cDNA, Taq DNA polymerase-mediated addition of 3´ A bases, unique cDNA with several truncated PCR products and unique cDNA containing multiple poly(A) sites [Liang and Pardee, 1992, Bauer, 1993, Callard et al, 1994, Averboukh et al, 1996]. Numerous methods to screen for false positives have been suggested, most of which involve subcloning and direct or reverse northern blotting of several clones for each cDNA tag selected [Li et al, 1994b, Hakvoort et al, 1994, Mou et al, 1994, Liu et al, 1996, Mathieu-Daude et al, 1996, Vogeli-Lange et al, 1996]. Recently, the RNase protection assay has also been combined with DDRT-PCR for this purpose [Wan et al, 1996]. Some groups also report the use of re-amplified PCR products for direct sequencing or the direct probing of northern blots as a screening strategy [Callard et al, 1994, Reeves et al, 1995, Wang and Feuerstein, 1995]. The latter methods are not advisable because the re-amplified PCR products themselves may be false positives and may consist of overlapping bands [Debouck, 1995]. Other more basic methods include running out duplicate, triplicate and even quadruplicate RNA samples and repeating the individual reactions that
appear to give differential bands. Also the use of native rather than denaturing gels helps to reduce banding complexity.

Non-specific background “noise” can also impair the useful analysis of banding patterns and several suggestions to overcome this problem have been put forward. One way is to replace the anchor primer in the PCR step with a second AP sequence. This avoids any misalignment of the oligo(dT) region and priming at internal poly(A) sites [Haag and Raman, 1994]. Another method of increasing primer specificity is to use longer sequences and alter PCR conditions to increase stringency [Linskens et al, 1995]. Other modifications include the use of different isotopes to label dNTPs, radio-labelling one of the primer groups, decreasing primer degeneracy and the use of modified bases in primer sequences [Hadman et al, 1995, Trentmann et al, 1995, Toshiham and Takeda, 1995, Zao et al, 1995, Liang et al, 1994, Rohrwild et al, 1995].

The final major problem associated with DDRT-PCR is selection of cDNA sequence tags in the 3' untranslated region (UTR). This arises because priming is targeted to the region adjacent to the poly(A) tail and as the sequence tags obtained are commonly in the order of 100-500 bp in length, the majority of products correspond to 3' UTRs. Problems arise since i) this region of the mRNA is frequently not included in sequence databases, ii) it shows great inter-organism variation and iii) in contrast to their coding regions homologous genes may demonstrate dramatic differences in their 3' UTR sequences. This necessitates the use of full length cDNAs from screening libraries or from PCR-generated techniques to obtain useful sequence information. An alternative strategy to determine 5' sequence information is the modification of PCR conditions to produce longer cDNA tags by increasing dNTP concentration and elongation steps [Averboukh et al, 1996]. Another suggestion has been the addition of an inverse-PCR step by making a series of primers from the known 5' end of the cDNA tag and “walking” towards the 5' end of the mRNA [Sompayrac et al, 1995].

Other modifications to the basic technique of DDRT-PCR have been developed to increase safety and for use in the automation of the process. These include the use of fluorescent-labelled primers for use in automatic DNA sequencers [Ito et al, 1995], silver staining of PCR products [Lohmann et al, 1995] and the development of sequencing machines specifically for use with DDRT-PCR generated cDNA tags [Averboukh et al, 1996, Luehrsen et al, 1997].
2.3 Applications of DDRT

The technique can be applied to a variety of systems, both simple and complex. The most basic level of application is that of a single cell line at different levels of confluence. Liang and Pardee's introductory DDRT-PCR paper compared gene expression in quiescent versus cycling mouse A31 cells [Liang and Pardee, 1992]. The next group of applications examine cell cultures in the presence/absence of treatment with various agents such as growth factors, ionising radiation and differentiation-inducing chemicals. Such studies have used DDRT-PCR to investigate the effect of polypeptide mitogens on the cell cycle using fibroblast growth factor with NIH 3T3 cells [Hsu et al, 1993, Donohoe et al, 1994] and human endothelial cells [Li et al, 1994b], to examine mechanisms of cell damage in diabetes by identifying glucose-induced genes in aortic smooth muscle and retinal blood vessels [Nishio et al, 1994, Aello et al, 1994] and to evaluate phagocyte defence mechanisms with stress-inducible genes in Salmonella typhimurium [Wong and McClelland, 1994]. Other applications include the study of adipocyte differentiation [Edwards et al, 1994], retinoic acid gene induction and human myeloid cell differentiation [Burn et al, 1994], the effect of γ- radiation on cell lines to identify targets of p53 [Okamoto and Beach, 1994], and studies on steroid-induced gene expression [Chapman et al, 1995]. More recent applications have included elucidation of novel mRNA transcripts expressed by Legionella pneumophila during host infection [Kwaik and Pederson, 1996], studies on estrogen-regulated gene expression in human vasculature [Koike et al, 1996], the effect of various mitogens on human keratinocytes [Frank et al, 1997] and gene expression studies in activated CD4+ T-cells [Liu and Raghoothama, 1997].

A more complex application is the comparison of normal and tumour cell lines or tissues, with many studies having utilised the technique to discover genes associated with several neoplastic systems. Differential display of RNA from breast cancer cell lines [Liang and Pardee, 1992, Liang et al 1993, Sager et al, 1993] and from fresh tissue [Allred et al, 1993, Watson and Flemming, 1994] has revealed a number of sequence tags related to gene expression at various stages of the disease. Similar results have been achieved with ovarian cancer [Mok et al, 1994], epidermal cancer [Sun et al, 1995], gastric cancer [Salesiotis et al, 1995] and lung cancer [Chen et al, 1996].

Chen's group compared a normal lung fibroblast cell line with an epithelial lung tumour cell line and found 30 differentially expressed bands. Only one band showed differential
expression on northern blot analysis, with increased signal in the tumour RNA sample. Subsequent analysis of this cDNA tag, named N8, verified similar increased expression patterns in RNA from human lung tumour specimens and other human lung tumour cell lines. Screening of cDNA libraries demonstrated a 1.9 kb clone, containing an open reading frame and coding for 184 amino acids. Analysis of its structure and in vitro translation studies confirmed the sequence did encode a protein. Somatic cell hybridisation and fluorescent in situ hybridisation experiments revealed the sequence localised to chromosome 8q13. Expression studies using non-pulmonary human tissues and various animal tissues, both adult and fetal, indicated the message was widespread throughout the epithelial layers and was expressed during fetal development. The group did not, however, determine the identity of N8.

DDRT-PCR has allowed the study of elaborate in vivo systems such as tissue regeneration following hepatectomy [Bauer, 1993] or amputation [Simon and Oppenheimer, 1996] and aspects of chronic graft rejection [Russell et al, 1994, Utans et al, 1994]. The final level of current applications of DDRT-PCR is to identify genes expressed during normal embryonic development where quantities of tissue are limiting [Zimmerman and Schultze, 1994] and in growth regulatory pathways [Green and Besharse, 1996].

2.4 Applications of DDRT-PCR to lung research

The potential use of a technique like DDRT-PCR in lung research is huge! Isolated cell cultures can be used to address the effect of growth factors on differentiation, to examine changes in gene expression induced by hypoxic damage, radiation changes and potential pulmonary carcinogens. Matched tumour/normal tissues from lung cancer patients can provide information on novel tumour suppressor genes and oncogenes, as well as tumour progression and metastasis. Data on the regulation of multicellular development processes can be obtained with fetal lung organ cultures and lung bud cultures using DDRT-PCR. Complex in vivo processes such as normal lung development, injury and repair and tumourigenesis can also be studied. (For a complete review see Sunday, 1995).
2.5 The Future...

Since the development of DDRT-PCR, several new alternative methods for enriching differentially expressed genes have been produced. They are based on subtractive hybridisation techniques and include suppression subtractive hybridisation (SSH) [Diatchenko et al., 1996], representational difference analysis (RDA) [Edman et al., 1997] and subtractive hybridisation using magnetic oligo(dT) beads [Meszaros and Morton, 1996].

SSH combines subtractive hybridisation with a method for normalising the abundance of cDNAs within the target population, called suppression PCR [Siebert et al., 1995]. Briefly, a test population of cDNAs is hybridised against a driver cDNA population as previously, except this time both cDNA populations are restricted with a 4-base cutter enzyme, the test cDNAs are divided into 2 groups and each is ligated to a different double-stranded cDNA adaptor. Each of the 2 test cDNA-adaptor groups is then hybridised to the driver cDNA population, and the remaining single-stranded test cDNAs are mixed together. The ends of each adaptor group are filled in to create primer binding sites and the reaction is amplified by PCR. The principle of the process being that only unique, complementary cDNAs which are present in both test samples will hybridise, carrying both adaptor-primer groups necessary for exponential amplification during PCR. Unique cDNAs not present in both test groups will only possess a single adaptor-primer site and will not be amplified. Any hybridisation between cDNAs containing the same adaptor-primer group at both ends will result in self-annealing with the formation of a loop-structure that will not undergo amplification. This is the "suppression effect". See figure 2.3. A commercial kit for this method is now being marketed by Clontech, called PCR-Select™ cDNA Subtraction.

RDA was originally described by Lisitsyn et al. [1993] for the analysis of differences between genomic DNA samples and has since been modified by several groups for use with cDNA [Edman et al., 1997]. The principles of cDNA-RDA are very similar to those of SSH. Tester and driver cDNA pools are cleaved with restriction enzymes to form “amplicons”, these are ligated to oligonucleotide adaptors and amplified by PCR. The adaptors are then removed by cleavage and new adaptors are ligated to the 5' ends of the tester cDNAs only. The tester and driver populations are mixed and a hybridisation-amplification step is performed. This selectively enriches the sample for double-stranded tester cDNAs as only these hybrids carry primer annealing sites (from the 5' adaptor) which will allow exponential amplification.
1. Test cDNA is synthesised from mRNA. Test cDNA is digested with restriction enzymes and divided into 2 portions. Reference cDNA (driver) is also treated similarly. Each portion of test cDNA is ligated with a different cDNA adaptor molecule.

2. During the first hybridisation, an excess of driver cDNA is added to each test sample. Molecules $a-d$ are generated. Single-stranded type $a$ molecules are enriched for differentially expressed sequences, as common cDNAs will form type $c$ molecules with the driver cDNA. During the second hybridisation, the 2 original portions of tester cDNA are mixed together. Only the type $a$ molecules are able to re-associate to produce type $e$ hybrids.

3. These new molecules now possess different adaptor sequences at either end. Only these molecules will undergo exponential amplification using PCR primers for the adaptor-ends.
Figure 2.3  Schematic diagram of the principles of suppression subtractive hybridisation.

1. Test cDNA + adaptor 1  
   \[\text{Driver cDNA (in xs)}\]  
   Test cDNA + adaptor 2  
   \[\text{First hybridisation}\]

2. a  
   b  
   c  
   d  
   +  
   +  

3. a, b, c, d + e  
   \[\text{Second hybridisation}\]  
   Mix samples, add fresh denatured driver & anneal

   a, b, c, d + e

   \[\text{Add primers} + +\]  
   Fill in ends

   a  
   b  
   c  
   d  
   +  
   +  
   e  

   \[\text{Amplify by PCR}\]

   a & d no amplification
   b & b' no amplification
   c linear amplification only
   e exponential amplification

Redrawn from Diatchenko, 1996 (see References)
Single-stranded products are removed by digestion with mung bean nuclease, leaving double-stranded products only. These then undergo a second round of enrichment, beginning with restriction enzyme digestion. The driver cDNA acts as a competitive inhibitor to the self-annealing of the tester cDNAs common to both cDNA pools. Unique target cDNAs will be present only in the tester pool and will be enriched relative to the other tester cDNAs following PCR. Two or three rounds of target enrichment are usual, representing an increase of $10^5$ in target amplicons [Lisitsyn et al, 1993].

The third subtractive hybridisation-based method involves the capture of driver mRNA onto magnetic beads coated with oligo(dT). Following hybridisation with tester cDNA, double-stranded mRNA/cDNA hybrids remain stuck to the beads and can easily be removed. Single-stranded tester cDNA is left in solution and can then be tailed with oligo(dA) and annealed to a new batch of beads. Second strand synthesis and PCR amplification are then performed, the PCR product acts as a template for the generation of probes with which to screen the tester cDNA library. This method has the advantage of providing a solid substrate to support the double-stranded hybrid molecules, allowing their removal from the reaction without the loss of low abundance single-stranded tester molecules that commonly occurs with chromatography and biotin labelling separation methods [Meszaros and Morton, 1996].

These methods have several advantages over DDRT-PCR. For example, they minimise the loss of low abundance mRNAs, with transcripts as low as 0.001% of the poly(A) pool being detected [Travis and Sutcliffe, 1988]. There is an increased likelihood of detecting fragments within open reading frames and the number of false-positives and background signal is diminished [Edman et al, 1997]. The one disadvantage of all these techniques is that they require large amounts of starting material. SSH needs 2 µg poly(A) RNA, RDA uses 20 µg total RNA per reaction and the magnetic bead method requires a massive 2 mg of total RNA.

Anyone considering future projects involving differential gene expression would do well to consider these pros and cons carefully when selecting an appropriate protocol. Nevertheless, DDRT-PCR remains a valid choice for experiments where the sizes of tissue/cell samples are limited and a recent conference on DDRT-PCR methodology at Cold Spring Harbor attracted more than 400 participants from all over the world, confirming that DDRT-PCR remains a very popular choice [Sager, 1997]!
Chapter 3 -

Methods
Chapter 3 Methods

(Please also refer to Appendix I for material sources)

3.1 Sample collection

3.1.1 Cell culture

The following human cell lines were used; L-132 (embryonic normal lung epithelium), A-549 (Clara cell adenocarcinoma), SKLU-1 (pulmonary adenocarcinoma) and MOR-P (pulmonary adenocarcinoma). Lines L-132 and SKLU-1 were purchased from ECACC, CAMR, UK. Lines A-549 and MOR-P were gifts. Cultures of L-132 were grown in Medium 199, while the other lines were grown in RPMI medium. All media were supplemented with 10% fetal calf serum in the absence of antibiotics and the cultures were maintained at 37°C with added 5% CO₂. Cultures were split at confluence via 1:4 seeding using trypsin/EDTA. All cultures were subjected to the same conditions of growth and were harvested at the same level of confluence. All cultures were free from mycoplasma contamination. Cell pellets of $10^7$ cells/ml from each cell line were prepared for RNA extraction.

3.1.2 Human tissue collection

Five matched cases of pulmonary adenocarcinoma were studied. The only criterion for selection was the availability of sufficient tissue for RNA preparation. Specimens were received fresh in the laboratory within 30 minutes of removal from the patient, 1cm³ tissue pieces were snap frozen in liquid nitrogen or isopentane and stored in vapour-phase liquid nitrogen until required for RNA extraction. In each case samples of tissue were taken from the area of the tumour, the area immediately adjacent to the tumour and from macroscopically uninvolved tissue distal to the tumour. Part of each sample was also fixed in 10% formol saline for 24-48 hours and embedded in paraffin wax for diagnostic use. These fixed sections were cut and stained with haematoxylin and eosin for tissue diagnosis by a consultant histopathologist.

All samples were obtained with the approval of the Local Ethical Committee and patients were provided with additional consent forms outlining the aims of the project. A summary of human tissue stocks collected is given in table 3.1.
Table 3.1. Summary of human tissue samples collected

<table>
<thead>
<tr>
<th>Tissue type:</th>
<th>Histological type:</th>
<th>Number matched/unmatched:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lung / AC</strong></td>
<td>AC</td>
<td>5 matched</td>
</tr>
<tr>
<td></td>
<td>AC</td>
<td>10 unmatched</td>
</tr>
<tr>
<td><strong>Lung / other tumours</strong></td>
<td>Carcinoid</td>
<td>1 unmatched</td>
</tr>
<tr>
<td></td>
<td>LCC</td>
<td>2 matched</td>
</tr>
<tr>
<td></td>
<td>Leiomyosarcoma (M)</td>
<td>1 matched</td>
</tr>
<tr>
<td></td>
<td>Rhabdomyosarcoma (M)</td>
<td>1 matched</td>
</tr>
<tr>
<td></td>
<td>SCC</td>
<td>4 matched</td>
</tr>
<tr>
<td></td>
<td>SCLC</td>
<td>1 matched</td>
</tr>
<tr>
<td><strong>Lung / non-neoplastic</strong></td>
<td>Bronchiectasis</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Bronchogenic cyst</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>COAD</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Inflammation</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Non-lung / AC</strong></td>
<td>Colonic AC</td>
<td>1 matched</td>
</tr>
<tr>
<td></td>
<td>Gastric AC</td>
<td>1 matched</td>
</tr>
<tr>
<td></td>
<td>Oesophageal AC</td>
<td>1 matched</td>
</tr>
<tr>
<td></td>
<td>Ovarian AC</td>
<td>1 unmatched</td>
</tr>
<tr>
<td></td>
<td>Pancreatic AC</td>
<td>1 matched</td>
</tr>
<tr>
<td><strong>Non-lung / non-neoplastic</strong></td>
<td>Cardiac muscle</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Cerebellum</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Colon</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Liver</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Stomach</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Key: AC = adenocarcinoma  COAD = chronic obstructive airways disease
LCC = large cell lung cancer  M = metastatic tumour
SCC = squamous cell lung cancer  SCLC = small cell lung cancer
N/A = not applicable
3.1.3 Animal tissue collection

The male A/J mice used in this study were purchased from Harlan Orlac, UK. The animals were brought in when 8 weeks old and left to acclimatise to their surroundings for 2 weeks before treatments commenced. The animals were cared for according to MRC animal facility guidelines. They were maintained in at an ambient temperature of ~21°C in plastic isolators with a filtered air supply, with 5 animals per box. The mice were given standard mouse diet RM-1 feed and tap water as required. They were exposed to 8 hour cycles of light and 16 hours of darkness. Tumours were induced in thirty mice by a single intraperitoneal (i/p) injection of 10% urethane solution (100 mg/ml) at 10 μl/g body weight. Equal numbers of these animals were sacrificed at 3, 6 and 9 months following treatment. Twenty control mice were dosed with an equivalent volume of 0.9% saline solution and allowed to develop spontaneous lung tumours over a 9 month period. All animals were sacrificed by lethal i/p injection of sodium pentobarbitone (200 mg/ml) at 200 μl per 30g body weight. The lungs and other organs were removed immediately. The tissue was then either snap frozen in liquid nitrogen for use in RNA extraction or fixed in 4% paraformaldehyde solution at 4°C overnight before being embedded in paraffin wax. These tissue blocks were cut and mounted on silane coated slides for use with in situ hybridisation.
3.2 RNA preparation

Extraction of total RNA was achieved using either method i) or ii), below, and both methods are based on similar principles [Chirgwin et al, 1979, Chomczynski and Sacchi, 1987]. The sample is first homogenised in a denaturing solution containing two potent inhibitors of RNase, guanidinium isothiocyanate (GITC) and β-mercaptoethanol. GITC acts with N-lauryl sarcosine (SDS) to disrupt cell membranes and nucleoprotein complexes, thus allowing RNA to be released into a solution free of proteins. The RNA is then isolated by extraction in acidic phenol-chloroform, the low pH causes the RNA to remain in the aqueous phase while DNA and proteins are partitioned in the organic phase. Once isolated, the RNA can be precipitated with isopropanol. A second wash with GITC and re-precipitation in isopropanol removes any remaining protein contaminants. Since GITC, phenol and β-mercaptoethanol are all hazardous chemicals, the advantage of RNAzol™ B is that it comes ready to use and so minimises contact with these reagents.

3.2.1 Extraction of RNA using RNAzol™ B protocol

Approximately 100 mg of tissue or 1 x 10^7 cells were homogenised with 2 ml of RNAzol™ B reagent. Two hundred μl of chloroform was added and mixed thoroughly, the samples were then left on ice for 5 minutes before centrifugation at 13,500 rpm for 15 minutes at 4°C. The aqueous phase was transferred to a clean tube, where an equal volume of cold isopropanol was added and the samples were left at 4°C for 15 minutes. Following centrifugation as before, the supernatant was removed and the resulting RNA pellet washed in 75% ethanol and re-centrifuged at 6,500 rpm for 8 minutes at 4°C. The pellet was vacuum-dried for approximately 10 minutes and resuspended in 20-50 μl of DEPC.H_2O containing RNase inhibitor. Samples were stored at -80°C.
3.2.2 Extraction of RNA using modified acid GITC protocol

Similar quantities of starting material to those stated in section 3.2.1 were added to 1 ml of GITC/β-mercaptoethanol denaturing solution (solution D). See Appendix III. Solution D contained 7.2 μl of β-mercaptoethanol for every 1 ml GITC. The sample was split into aliquots of 500 μl and to each was added the following; 50 μl 2M sodium acetate (pH 4.0), 500 μl aqueous phenol and 100 μl chloroform. After mixing and centrifugation as above, the aqueous phase was removed and added to an equal volume of cold isopropanol and re-centrifuged. One hundred μl of solution D and cold isopropanol were added to the RNA pellet and the mixture was left at -20°C for 15 minutes. Following centrifugation, the supernatant was removed and the pellet washed, re-spun, dried and resuspended as above.

3.2.3 RNA purification

In order to remove any traces of contaminating DNA all RNA samples were DNase treated before use in differential display. Up to 50 μl of RNA solution was added to a DNase digest containing the following; 6 μl 10 x DNase buffer, 2.5 μl human placental RNase inhibitor (100 U) and 1 μl DNase (RNase-free) (10 U/μl) to a final volume of 60 μl. The reaction mixture was incubated at 37°C for 30 minutes. One phenol-chloroform extraction was performed to remove any traces of protein. The aqueous layer was removed to a clean tube and the RNA precipitated with 0.1 volume 3M NaOAc (pH5.2) and 2 volumes cold 100% ethanol at -80°C for a minimum of 1 hour. The samples were centrifuged at 13,500 rpm for 10 minutes at 4°C, the RNA pellet was washed in 70% DEPC.ethanol and resuspended in 20 μl DEPC.H2O. Optical density measurements were made via spectrophotometry, using a Perkin Elmer Cetus UV/VIS Lambda 2 scanning spectrophotometer, and the concentration of RNA was calculated. Five μg RNA were run out on a 2% agarose gel to assess the purity and check the concentration of each sample.
3.2.4 RNA quality control

In order to assess the ability of the RNA to be converted into cDNA and amplified, all purified RNA samples were subjected to reverse transcription PCR RT-PCR before use in differential display. *K-ras* primers were selected for this purpose as the *K-ras* gene has a high expression rate in all tissues [Leon et al, 1987, Hesketh, 1994]. Any sample that failed to amplify with these primers would, therefore, be unlikely to be amplified using primers for less commonly expressed genes and would not be suitable for use in DDRT-PCR. The primer pair was selected to amplify across intron 1 of *K-ras* and act as an internal control for the presence of genomic DNA contamination. The PCR product from cDNA amplification should be a single band of ~290 base pairs. Genomic DNA amplification spans an intron in excess of 12,500 base pairs in length [McGrath, 1983]. This size of template cannot easily be amplified within the current parameters of PCR. Consequently, multiple bands of varying lengths are produced which represent random anealing of the primers. See Appendix II for primer sequences.

First strand cDNA synthesis was performed as follows; 10 μl total RNA (1 μg), 1μl T_{12}MN anchor primer (10 μM), 4μl 5 x RT buffer, 2 μl DTT (0.1 M), 1 μl mixed dNTPs (250 μM) in a final volume of 20 μl were mixed and incubated at 65° C for 5 minutes to denature the samples. The anchor primers annealed at 37 °C for 10 minutes before 1 μl M-MLV reverse transcriptase (200 U) was added. The mixture was left at this temperature for a further 50 minutes, following which the enzyme was heat inactivated at 95° C for 5 minutes to terminate the reaction. The samples were then centrifuged briefly and stored at -20° C until required. Amplification of the first strand cDNA was performed as follows; a master mix containing 5 μl 10 x PCR buffer, 1 μl mixed dNTPs (10 mM), 1 μl *K-ras* primer #1 and #2 (50 μM each) and 0.2 μl *Taq* polymerase (5 U/μl) in a final volume of 49 μl per tube was mixed and left on ice. One μl of cDNA solution was added to each aliquot of the master mix and overlaid with 50 μl mineral oil. The reaction tubes were placed in a Perkin Elmer Cetus 480 Thermal Cycler programmed for 94° C for 2 minutes followed by 40 cycles of 94° C for 1 minute, 50° C for 1 minute and 72° C for 1.5 minutes. On completion, 9 μl of each sample were run out on a 2% agarose gel and visualised by ethidium bromide staining. Only those
samples which had been successfully amplified by the K-ras primers were used in differential display reactions.

3.2.5 mRNA extraction

Messenger RNA was also extracted from cell pellets for use in Northern blotting hybridisation using the Dynabead protocol. This method uses a denaturation solution to disrupt cell components and oligo(dT)-coated magnetic beads to bind poly(A) RNA. The poly(A) RNA is then washed off the beads and stored at -80 °C until required.

3.3 Precautions for differential display RT-PCR

3.3.1 Treatment of equipment

Due to the extreme sensitivity of DDRT-PCR, even trace amounts of contaminating DNA from any source can lead to the production of anomalous banding patterns. Various precautions were employed to ensure all equipment was as clean as possible;

a] all reactions were set up in class 2 safety cabinet cleaned with Virkon® disinfectant and swabbed with 100% ethanol,

b] a set of pipettes, dedicated for use only with DDRT-PCR, was obtained and cleaned with Virkon® and absolute ethanol before and after use,

c] sterile, aerosol resistant, RNase-free tips and sterile, RNase-free tubes were used for all reactions,

d] tips and tubes were UV-crosslinked for 15 minutes prior to each experiment,

d] sterile, powderless gloves were used during all reactions.

3.3.2 Treatment of reagents

All reagents were aliquoted on arrival or immediately following assembly, placed in clean sealed boxes and stored at -20 °C until required. Partially used aliquots were neither re-frozen nor re-used at any time.
3.3.3 **Negative controls for DDRT-PCR**

To ensure no DNA contamination was present in any of the reagents used in DDRT-PCR, several negative control samples were set up with each new batch of reagents used. For cDNA synthesis controls, each anchor primer mix was set up using water instead of RNA and these water blanks were then placed into the PCR step. A tube containing reagent mix and RNA but without M-MLV enzyme was also prepared. Similar negative controls for the PCR components were also set up. Subsequent treatment of these blanks was the same as the test samples.

3.4 **Differential display RT-PCR**

DDRT-PCR was carried out essentially according to the method described by Liang and Pardee [1992] using anchor primer sequences from the RNA Map Kit A™ protocol. The principles of this technique are explained in Chapter 2.

3.4.1 **cDNA synthesis by reverse transcription**

For each RNA sample four RT reactions were prepared using each of the four anchor primers T12MA, MC, MG and MT as follows: 4 μl 5 x RT buffer, 2 μl DTT (0.1 M), 1.6 μl dNTP mix (250 μM), 2 μl T12MN primer (10 μM), 2 μl total RNA (DNA-free) (0.1 μg/μl) in a final volume of 20 μl were mixed and incubated at 65° C for 5 minutes to denature the samples. The reaction was left to anneal at 37° C for 10 minutes, before 1 μl M-MLV reverse transcriptase (200 U) was added to each tube. cDNA synthesis continued for a further 50 minutes at this temperature before termination by heating to 95° C for 5 minutes. At the end of the reaction the tubes were briefly centrifuged and stored at -20° C for later use.
3.4.2 cDNA amplification by PCR

Master mixes containing the following were prepared; 2.0 μl 10 x PCR buffer, 1.6 μl dNTP mix (25 μM), 2 μl arbitrary primer (2 μM), 2 μl Ti2MN primer (10 μM), 1 μl α-32P- dATP (1200 Ci/mmole) and 0.2 μl Taq polymerase (5 U/μl) in a final volume of 18 μl per reaction were mixed, aliquoted and kept on ice. Two μl of RT-mix (containing the same Ti2MN as present in the PCR mix) was added to the appropriate aliquot of PCR mix and overlaid with 50 μl mineral oil. The reaction tubes were placed in a Perkin Elmer Cetus 480 Thermal Cycler programmed for 40 cycles of 94° C for 30 seconds, 40° C for 2 minutes and 72° C for 30 seconds followed by a 10 minute incubation at 72° C.

3.4.3 Separation and display of PCR products

Following DDRT-PCR, samples were run out on a 6% polyacrylamide non-denaturing gel containing 40% acrylamide/bis (19:1), 5 x TBE, 10% ammonium persulphate and TEMED and poured between glass sequencing plates, prepared according to the protocol described in Maniatis [Sambrook et al, 1989]. It should be noted that polyacrylamide gels are toxic, very fragile and have a tendency to “self-destruct” at the slightest provocation. They should, therefore, be treated with the utmost care at all times!

The gel was left at 4° C for at least 2 hours to polymerise before use. Three μl of PCR product were mixed with 2 μl of loading dye and loaded into the sequencing gel. The gel was electrophoresed at 50-60 Watts/1800 V for approximately 3 hours until the xylene dye had travelled to within 10 cm of the bottom of the gel. The gel assembly was dismantled, the gel transferred onto 3MM paper and covered with Saran™ wrap. The gel was then vacuum dried at 80° C for 90 minutes, following which the Saran™ wrap was removed and the gel placed into an autoradiograph cassette. The gel was orientated with fluorescent tape (Tracker tape™, Amersham), before being covered by a sheet of Fuji-RX autoradiography film and left at room temperature for 1-3 days before being developed.
3.5 Extraction and re-amplification of bands

3.5.1 Extraction of bands from the sequencing gel

Selected bands were orientated on the gel using the fluorescent tape. Every band chosen for analysis was marked at each corner by punching holes through the autoradiograph film with a needle. The bands were then cut out using a sterile scalpel blade, placed in sterile microtubes and soaked in 100 μl dH₂O. The remainder of the gel was wrapped in Saran™ film and stored between two sheets of 3MM paper. The bands were boiled for 15 minutes and spun at 13,500 rpm for 2 minutes to pellet the debris. The supernatant was removed to a clean tube and the DNA precipitated by the addition of 10 μl 3M NaOAc (pH 5.2) and 450 μl 100% ethanol, with 5 μl glycogen (10 mg/ml) added as a carrier. The solution was left at -80° C for a minimum of 30 minutes. The tubes were then centrifuged at 13,500 rpm for 10 minutes at 4° C to pellet the DNA. The supernatant was removed and the pellet washed in 85% ethanol before being resuspended in 10 μl dH₂O. The DNA was stored at -20° C until use.

3.5.2 Re-amplification of cDNA from bands

Re-amplification of the DNA in the bands was done using the same primer combination and conditions as used in the original PCR step, except that this time no isotope was added and the dNTP mix was at 250 μM starting concentration. Master mixes were prepared as previously described for a final volume of 36 μl per reaction. Four μl of cDNA solution was added to each tube and the reactions overlaid with mineral oil. The tubes were placed in a thermal cycler as before and subjected to the same cycling program. Thirty μl of re-amplified PCR product was run out on a 1.5% agarose gel and stained with ethidium bromide. Once confirmation of success was obtained, the PCR product was extracted from the gel with a QIAEX purification kit and the cDNA solution was then used in a subcloning step.
3.6 Subcloning of template DNA

The basic principles of subcloning are simple; a DNA fragment (insert) is transferred into a specialised carrier (vector) which is then introduced into bacterial host cells. There are 4 steps to this process. Firstly, the insert and vector DNAs are cut with suitable restriction enzymes to produce compatible ends which can stick together when treated with T4 DNA ligase (ligation step), the ligated vector-insert molecules are then introduced into specially treated host cells (transformation of competent cells), and lastly, screening procedures are employed to select recombinant transformants. Once recombinants have been produced, they provide an unlimited source of insert DNA which can be used for a variety of procedures, such as sequencing and probe preparation.

The TA Cloning® kit was used for the ligation and subcloning of differential display bands. The details are presented below. See figure 3.1 for details of the cloning vector used.

3.6.1 Ligation

The TA Cloning® method utilises the non-template-dependent activity of Taq polymerase in adding single A bases to the ends of PCR products. This allows efficient ligation into linearised pCR™II vector which contains single overhanging T's, hence the name “TA” cloning. For optimal ligation efficiency the PCR products must be less than 24 hours old, as the A-overhangs are removed over time. A ligation mix containing the following was prepared in a final volume of 10 µl: 1-3 µl fresh PCR product, 1 µl 10 x ligation buffer, 2 µl pCR™II vector (25 ng/µl), 1 µl T4 DNA ligase (4 U/µl). The reaction was left at 14° C for a minimum of 4 hours. The reaction was stored at -20° C until required for the transformation reaction.

3.6.2 Transformation and screening

Two µl of ligation mix was added to 50 µl competent INVαF cells and left on ice for two minutes. The cells were heat shocked for 30 seconds at 42° C and left on ice for a further 2 minutes before 450 µl SOC medium was added. The reactions were incubated at 37° C for 1 hour at 225 rpm in a horizontal rotary shaker before being stored on ice. This step allows time
Figure 3.1 Map of cloning vector pCR™II

The sequence detailed above represents the pCR™II vector sequence with a PCR® product inserted by TA Cloning®. Note that the pCR™II vector sequence list in the following pages is modified at the unique EcoRI site during preparation for TA Cloning® so the inserted PCR product is flanked on each side by EcoRI sites, as shown above.

Arrows (↑) indicate the start of transcription for Sp6 and T7 RNA polymerases, respectively.

Nsi I restriction sites are indicated as this enzyme was used to remove the insert from the vector DNA.

for the expression of the ampicillin resistance gene in the transformants. Two hundred μl from each transformation mix was spread onto L-agar plates containing 50 μg/ml ampicillin and 2% X-gal. The plates were inverted and placed in an incubator at 37° C overnight.

Selection of transformants containing the insert DNA is achieved firstly by the insertion of an antibiotic resistance gene [Bolivar et al, 1977], in this case for ampicillin, and secondly by blue-white screening [Yanisch-Peron et al, 1985]. Antibiotic resistance is fairly self-explanatory. The host cells are naturally sensitive to ampicillin and only those containing vector molecules, which are ampicillin resistant, will be able to survive in medium supplemented by antibiotic. Blue-white screening involves insertional inactivation of a specific gene. X-gal is a chromogenic substrate for the enzyme β-galactosidase and when cleaved by the enzyme produces a blue compound. Partially defective copies of the β-galactosidase gene are present in the cloning vector and in the host cell, and the two genes will complement each other to produce a functional β-galactosidase that will metabolise X-gal when the host cell contains a vector molecule. Insert DNA inactivates the gene and no functional β-galactosidase is produced and no blue product is formed. Host cells containing vector without insert DNA, therefore, form blue colonies on agar plates, while those containing insert DNA form white colonies. In order to maximise colour development and optimise colony identification, the plates were left at 4° C for 2 hours following the overnight incubation.

3.6.3 Plasmid preparation and purification of DNA

Most plasmid "minipreps" are variations on the alkaline lysis method described by Birnboim and Doly [1979]. The principles of this technique are as follows. Host bacterial cells are lysed in NaOH and SDS to allow the plasmid DNA to escape. The solution is then neutralised with KOAc, causing the selective precipitation of bacterial chromosomal DNA. This becomes trapped in the cell debris following centrifugation, allowing the plasmid DNA to be retained in the supernatant. The plasmid DNA can then be precipitated with isopropanol and purified from any residual contaminants by banding on a CsCl/ethidium bromide gradient [Joly, 1996].
Several white colonies from each plate were lifted using a sterile toothpick and placed into separate 5 ml aliquots of LB broth/ampicillin for an overnight incubation as described above. Three ml of each culture was centrifuged at 13,500 rpm for 3 minutes and the pellet resuspended in 100 μl GTE (50 mM glucose, 25 mM Tris-HCl pH 8.0, 10 mM EDTA). To this was added 200 μl 0.2 M NaOH/0.1% SDS solution and the pellet was shaken vigorously to ensure complete lysis. One hundred and fifty μl KA solution (3 M KOAc, 2 M acetic acid) was added, the suspension was mixed well and centrifuged for 10 minutes as before. After removal of the supernatant, nucleic acid precipitation was achieved with isopropanol. The pellet was spun down, washed with 70% ethanol and air dried for 2 minutes. The DNA was resuspended in 30 μl TE to which 70 μl CsCl (1 g/700 μl) and 10 μl ethidium bromide (10 mg/ml) were added. The tubes were shaken and centrifuged for 5 minutes. The upper layer was removed and 50 μl isopropanol added and mixed. This step was repeated until the lower layer was clear. The CsCl was removed by the addition of 400 μl TE, 50 μl 3 M NaOAc and 750 μl isopropanol. The samples were then centrifuged for 10 minutes, the pellet washed in 7% ethanol and air dried. The DNA was finally resuspended in 20 μl Tris (pH 8.0) and stored at 4°C until use.
3.7 Sequencing of template DNA

3.7.1 Manual sequencing of template DNA

DNA sequencing was performed according to the Sanger dideoxy chain termination method using a Sequenase™ 2.0 sequencing kit [Sanger et al, 1977]. This method uses base-specific dideoxynucleotides (ddNTPs). The ddNTPs are analogues of the normal dNTPs but differ from dNTPs in that they lack a hydroxyl group at the 2' and 3' carbon atoms. The lack of the 3'-OH group prevents the addition of a new nucleotide to the chain, as no phosphodiester bond can be formed at this position, and the nucleotide chain is abruptly halted. The template DNA must first be made single-stranded either by alkali or heat denaturation. A complementary sequencing primer (~16-20 bases long) is used to bind to the flanking region of the template DNA on one strand. Reactions are set up in parallel using four dNTP mixes per template, each of which also contains a small amount of a different ddNTP (i.e. ddATP, ddCTP, ddGTP or ddTTP). Chain termination occurs randomly when ddNTP is incorporated into the dNTP chain at varying distances into the template sequence. This produces four groups of partial reactions, each containing terminated chains of differing lengths, with identical 5' ends (sequencing primer site) and variable 3' ends (ultimately ending with one of the ddNTPs). These partial DNA fragments can then be separated by denaturing gel electrophoresis using a polyacrylamide/bis gel containing urea. Polyacrylamide/bis forms a "molecular sieve" through which the DNA fragments must pass, while urea acts as a denaturing agent to maintain the fragments in single-stranded form by preventing hydrogen bonding between base pairs. Such gels have the ability to separate DNA fragments that differ by only a single nucleotide. The terminated dNTP mixes are then loaded separately into the gel. Radioactive nucleotide is incorporated into the termination mix so that the signals can be visualised by autoradiography. After the gel has been run, fixed and dried down, it is exposed to autoradiography film for several days before being read. The sequence is read from the base of the film upwards. Approximately 200-300 bases of sequence can usually be read from one gel. Sequence data can be increased if multiple reactions are prepared simultaneously and run on the same gel at staggered intervals, i.e. 2 hours, 4 hours & 6 hours. Overlapping sequences can then be linked together. See figure 3.2.
Figure 3.2. Dideoxy DNA sequencing reaction.

i) Single-stranded DNA template anneals to sequencing primer

\[ 5' \cdots CGACGGTAACGTCAT\cdots 3' \]  
\[ \leftrightarrow \]  
\[ \text{New DNA synthesis} \quad \text{Primer} \]

ii) Termination mixes are prepared with ddATP, ddCTP, ddGTP & ddTTP  
(only ddTTP mix shown)

\[ TA \quad n + 2 \]
\[ TGCAGTA \quad n + 7 \]
\[ TTGCAGTA \quad n + 8 \]
\[ TGCCATTGAGTA \quad n + 13 \]

Chain elongation stops once ddTTP is incorporated, resulting in DNA fragments of varying lengths. DNA fragments will be complementary to the template sequence.

iii) Termination mixes are loaded onto denaturing gel for resolution of products

Read gel autoradiograph starting from the bottom: +1 \(\rightarrow\) +15 (ATGACGTT...)
Both strands of the template DNA were sequenced using the T7 and M13-reverse primers. Terminated sequencing reactions were run out on a 6% denaturing polyacrylamide gel, prepared as previously described, with the exception that the gel mix contained 7 M urea. The dried gel was covered with a sheet of Fuji-RX autoradiography film and exposed overnight at room temperature before being developed. The sequences were read manually, formatted in the G(C)G style and checked against GenBank and EMBL databases for confirmation and identification via the BLAST and FAST nucleotide search programs.

3.7.2 Automated sequencing of template DNA

Template DNA was also sequenced using the automated Perkin Elmer ABI-PRISM Terminator™ Taq Primer cycle sequencing method with an ABI-377 sequencer in order to verify the manual readings and gain further sequence information. Cycle sequencing involves the use of a double-stranded DNA template which is effectively rendered single-stranded by linear PCR amplification. Unlike normal PCR, which uses a pair of primers to produce exponential amplification of the template, only a single primer is used here and ddNTPs are present in the dNTP mix. Consequently, product amplification is linear, thus giving single-stranded, chain-terminated fragments as in i). The method also utilises the incorporation of fluorescently-labelled ddNTPs into the PCR extension products. These are then run out on a denaturing gel which is then scanned by a laser and the sequence pattern is recorded automatically. This method also has the advantage of giving longer amounts of sequence information than would be obtained from conventional manual methods.

As above, both strands were sequenced with T7 and M13-reverse primers. One-2.5 μl template DNA (300 ng/μl) were added to 8.0 μl Terminator™ reagent mix and 3.5 pmol of sequencing primer (0.5 pmol/μl) in a final volume of 20 μl. The reaction mix was overlaid with 50 μl mineral oil and placed in a Perkin Elmer Cetus 480 Thermal Cycler programmed for 25 cycles of 96 °C for 30 seconds, 50 °C for 15 seconds and 60 °C for 4 minutes. On completion, the amplified reaction mix was added to 95 μl 95% ethanol and 2 μl 3M NaOAc and left on ice for 10 minutes to allow DNA precipitation. The DNA was pelleted by centrifugation at 13,500 rpm for 30 minutes at room temperature, washed with 70% ethanol,
air-dried and placed in an automated DNA sequencer. Sequence information was formatted and scanned against the databases, as before.

### 3.7.3 M13 single-stranded sequencing

Sequences were also verified using single-stranded sequencing in the M13 vector. This method was used for sequences that proved refractory to conventional double-stranded sequencing protocols. M13 is a single-stranded filamentous phage. It becomes transiently double-stranded, M13 replicative form (RF), during its infection of sensitive, F* E.coli host cells. At the end of the replicative phase M13 is released as single-stranded phage particles which continue the infective cycle. See figure 3.3.

#### 3.7.3.1 Preparation of M13 vector DNA

The M13 DNA must first be treated with an appropriate restriction enzyme and to create suitable termini for ligation to the insert. Then it is dephosphorylated with calf intestinal alkaline phosphatase (CIAP) to prevent re-circularisation during ligation. Approximately 2 μg M13 mp18 RF DNA were digested with restriction enzymes EcoRI or HindIII (in separate reactions) using 5 μl 10 x One-Phor-All buffer (Pharmacia) in a final volume of 50 μl, and left to incubate at 37 °C for 1 hour. Five μl of the digest was then run out on a 1% agarose gel against uncut phage to check the restriction efficiency. The restricted phage DNA was then added to 0.1 U CIAP (1 U/μl) and 10 μl 10 x One-Phor-All buffer in a final volume of 100 μl. The reaction was incubated for 15-30 minutes at 37 °C, after which the CIAP was heat inactivated at 85 °C for 15 minutes. The dephosphorylated, linear phage DNA was purified by phenol/ chloroform extraction and precipitated with 3M NaOAC (pH 7.0) and 100 % ethanol, as previously described above. The pellet was then rinsed in 70 % ethanol, air dried and resuspended in 10 μl ultrapure H₂O.

#### 3.7.3.2 Ligation of M13 with insert

The insert DNA was restricted with the same enzyme as the vector. The entire restriction digest was run out on a 2% agarose gel, the desired fragment was cut out with a
Figure 3.3 Single-stranded sequencing using M13 phage.

i) MCS
   Insert DNA within plasmid cloning vector
   M13 phage (DS) → Cut both at restriction site with enzyme E

ii) Ligate linearised phage with insert fragment

iii) Transform into host cells
    F+ host cell ➔ RF M13 (DS) ➔ ssM13/insert

iv) ssM13 phage particles break out of host cell
    Phenol/Chloroform ➔ PEG/NaCl ➔ Ethanol/Salt ppt. ➔ Clean ssM13/insert DNA ➔ SEQUENCING

KEY: DS = double-stranded, MCS = multiple cloning site
      ss = single-stranded, RF = replicative form of M13
scalpel and the DNA was purified using the QIAEX kit. The insert and vector DNA samples were quantified by running 2 μl on a 1.6% agarose gel before the ligation step. Ligation reactions were set up using 50 ng M13 DNA with 1μl 10 mM ATP, 1 μl 10 x ligation buffer, 2 μl T4 DNA ligase (1 U/μl) and a 3-9 molar excess of insert DNA in a final volume of 10 μl. The reaction was incubated at 16 °C overnight and then stored at -20 °C until required for the transformation.

3.7.3.3 Preparation of host cells

Bacterial JM105 stock cultures (Pharmacia) were grown in 1 ml of L-broth at 37 °C overnight before undergoing F* selection. The bacterial host cells must be a male strain containing the F factor. The M13 phage infects the host cells via the sex pili, which are coded for by the F factor. The F factor also carries genes vital for blue-white screening. Selection of F* cells was achieved by incubating the cultures on M9/ thiamine agar plates containing 25 μg/ml streptomycin at 37 °C for 16-72 hours. Individual colonies were then used to inoculate 5 ml aliquots of LB medium and were left to grow up overnight. The following day, fresh cultures were prepared with 300 μl of overnight culture in 40 ml of LB medium. Cells were incubated for 3.5 hours, or until the OD _{550} reached 0.4-0.5. The cells were then pelleted by centrifugation at 1000 g for 2 minutes and the supernatant was removed. The pellet was gently resuspended in 20 ml of ice cold 50 mM CaCl₂ and left on ice for 15 minutes. The cell suspension was then centrifuged as before, the pellet resuspended in 2 ml of ice cold 50 mM CaCl₂ and left on ice for a minimum of 15 minutes before being used for the transformation step. Competent cells prepared in this manner should ideally be used immediately to retain maximum transforming ability.

3.7.3.4 Transformation and infection

Two hundred μl aliquots of competent cells were mixed with 5-10 μl of ligation reaction and placed on ice for at least 40 minutes. The cells were then heat shocked at 42 °C for 1 minute and returned to ice. Before plating out, the following reagents were added to each M13 transformation reaction: 50 μl X-gal (2%), 10 μl IPTG (100 mM) and 200 μl (exponential growth) JM105 (F*) host cells. The entire contents were added to 3 ml of molten
0.7% agarose/LB medium (cooled to 50 °C) and poured quickly on to a dry L-agar plate. The plates were inverted and incubated at 37 °C overnight to allow plaque development. Host cells which have been infected by the phage particles grow more slowly than uninfected cells, appearing as plaques within the background bacterial lawn.

Unlike most plasmid vectors, M13 phage does not carry antibiotic resistance genes, so the selection of transformants relies entirely on blue-white screening. Phage containing insert DNA will produce white plaques, while plaques from those without an insert will be blue. Here, IPTG acts as an inducer of the enzyme β-galactosidase which cleaves X-gal to form a blue compound, as described previously in section 3.6.2. Several individual white plaques were then lifted using a sterile toothpick and used to inoculate separate host cell cultures, containing 50 µl JM105 (F') cells in 2 ml LB medium. The cultures were then incubated at 37 °C for 5 hours with vigorous shaking (>300 rpm) to ensure good aeration, thus enhancing phage growth.

### 3.7.3.5 Single-stranded DNA extraction

M13 phage carrying the DNA insert has now infected the host cells and released single-stranded phage particles into the culture. The RF phages, containing double-stranded DNA, will remain within the host cells and can be removed by centrifugation, leaving a supernatant containing only single-stranded phage particles. Polyethylene glycol (PEG) is used to precipitate these phage particles, and is followed by phenol-chloroform extractions to remove coat proteins, leaving intact single-stranded DNA containing the insert. The host cell pellet is retained for PCR analysis to confirm the presence of a correctly sized insert fragment.

The cultures were centrifuged for 5 minutes at 14,000 rpm to pellet the bacterial host cells. The supernatant from each sample was then split into 2 x 1 ml aliquots, one of which was frozen at -20 °C and the other was used for DNA extraction. Analysis of the DNA from the bacterial pellets was carried out first to determine if the clone contained insert DNA. This was achieved by resuspending the host cell pellet in 0.5 ml distilled H₂O, boiling it for 5 minutes and spinning down for 2 minutes to pellet the debris. Approximately 5 µl of this solution was used as the template in a PCR with M13 forward and reverse primers (1 pmol/µl final concentration) in a final volume of 50 µl. Cycling parameters for the amplification were as
follows; 95 °C x 30 seconds, 65 °C x 30 seconds, 72 °C x 30 seconds, for 30 cycles. An aliquot of PCR product was then examined for the presence of any insert by agarose gel electrophoresis. Once confirmation was obtained, single-stranded DNA extraction was performed. Two hundred µl of 20% PEG/2.5M NaCl was added to the phage-supernatant solution and mixed thoroughly. The reaction was left at room temperature for at least 15 minutes to allow the phage particles to collect in the PEG/NaCl. The precipitated phage was collected by centrifugation and all of the supernatant was carefully removed. The phage pellet was then resuspended in 100 µl of TE (pH 8.0) by vigorous vortexing and purified by phenol/chloroform extraction. The phage DNA was precipitated with 3M NaOAc (pH 5.2) and 100% ethanol, cleaned in 70% ethanol and finally resuspended in 20 µl Tris (pH 8.0), as previously described above. The resultant single-stranded DNA solution was then used for manual and automated sequencing using the M13 (-40) primer, as described in section 3.7, above.
3.8 Blotting of RNA and genomic DNA

3.8.1 RNA dot blots

To 20 μg aliquots of total RNA was added 4 μl DEPC.H2O and 6 μl RNA denaturing buffer (see Appendix III). The sample was heated to 65° C for 5 minutes then snap cooled on ice before 8 μl 20 x SSC was added. Nytran® membrane was wetted in 10 x SSC and allowed to dry slightly before 2 μl aliquots of the RNA solution were dotted on. The membrane was dried in air, UV cross-linked, wrapped in cling-film and stored at room temperature until hybridisation.

3.8.2 RNA northern blots

RNA must be kept in a denatured form to prevent the formation of complex secondary structures which impair mobility. The loading buffer and the gel contain agents to combat this problem [Darling and Bricknell, 1994]. Formaldehyde reacts with the amine groups of bases A, C and G to prevent the formation of base pairs and formamide disrupts hydrogen bonding, allowing the formaldehyde to react with the bases. The formamide must first be de-ionised to remove ionic contaminants which may hydrolyse RNA. MOPS is used in preference to Tris buffer because the formaldehyde would react with the amine groups in the Tris. (For MOPS buffer see Appendix III).

Thirty μg aliquots of total RNA in volumes of 5 μl were heated with 15 μl denaturing buffer (also containing 1 μl ethidium bromide and 2 μl loading buffer) and loaded onto a 1% denaturing agarose gel. The gel was electrophoresed at 20 V overnight (or 100 V for 3 hours) in 1 x MOPS buffer. The gel was then rinsed in DEPC.H2O several times to prevent contamination by formaldehyde, before being blotted onto Nytran® membrane over 3-5 hours using a capillary blotting apparatus with 10 x SSC. Completion of RNA transfer was determined by visualisation of both the gel and membrane under UV light. The membrane was rinsed briefly in 2 x SSC, blotted onto 3MM paper and crosslinked in a UV Stratalinker for 90 seconds. The membrane was wrapped in cling-film and stored at room temperature until hybridisation.

112
3.8.3 Southern blot of genomic DNA

Southern blotting allows the transfer of DNA fragments from an agarose gel onto a nylon membrane [Southern, 1975]. To aid transfer, DNA samples are first broken up into smaller pieces using restriction enzymes, before being sized fractionated by agarose gel electrophoresis (non-denaturing). The DNA in the gel is then denatured in NaOH and transferred to a membrane by capillary blotting for use in probe hybridisation studies.

Ten μg aliquots of genomic DNA prepared via phenol-chloroform extractions were first restricted in separate reactions with EcoRI and HindIII. Reactions were set at 37°C for 3 hours using 50 U enzyme in a final volume of 100 μl. Following heat inactivation at 65°C for 10 minutes, the volume of samples was reduced to 20 μl using phenol-chloroform extraction and precipitation with NaCl/ethanol. The samples were loaded onto a 1% agarose gel and electrophoresed for 3 hours at 100 V. The gel was washed once in denaturing solution (1.5 M NaCl, 0.5 M NaOH) for 30 minutes and then washed twice in neutralising solution (1.5 M NaCl, 0.5 M Tris-HCl pH 8) for 20 minutes, all washes were at room temperature. Capillary blotting and subsequent treatment of the gel and membrane are described in section 3.8.2, above.
3.9 Preparation of cDNA probes and hybridisation

3.9.1 cDNA probe preparation

Probe labeling was achieved by the random primer method described by Feinberg and Vogelstein [1983]. During this method, the template DNA fragment is heat denatured to allow random sequence oligonucleotides to anneal to each strand and the Klenow fragment of DNA polymerase I is then utilised to extend the strands. A single radiolabeled nucleotide, usually $\alpha^{32}$P-dCTP, is included in the dNTP mix to produce a uniformly radiolabelled double-stranded DNA probe. The Amersham Multiprime kit allows the rapid production of probes with activities > $1 \times 10^6$ cpm/ml, which is more than sufficient for most routine applications.

Fifty μl of frozen glycerol stock of plasmid containing the relevant insert was incubated overnight in 20 ml LB broth containing ampicillin. DNA was extracted as previously described and five μg were used in a restriction digest with 50 U NsiI to release the insert from the vector sequence. The restricted product was then purified from an agarose gel as described above. Random primer labelling with the Multiprime kit protocol was used to incorporate $\alpha^{32}$P-dCTP into the cDNA fragment. A reaction mix was prepared containing 25-100 ng denatured cDNA, 10 μl 5 x labelling buffer, 5 μl hexamers, 2.5 μl T12MN primer (10 μM), 5 μl $\alpha^{32}$P-dCTP and 2 μl Klenow fragment in a final volume of 50 μl and incubated at 37°C for 30 minutes. The reaction mix was purified through a NucTrap® Push column, the fraction eluted was heat denatured and snap cooled in ice for 2 minutes before being added to the membrane/hybridisation reaction.

3.9.2 Hybridisation conditions

The pre-hybridisation buffer contains agents to suppress non-specific binding, and thus reduces background signal. SDS acts as a wetting agent. Denhardt’s solution stops non-specific binding to the surface of the filter, and non-specific annealing to DNA is reduced by the inclusion of denatured, fragmented salmon sperm DNA. The addition of 10% dextran sulphate reduces the effective reaction volume by excluding the probe from the solution and thus increasing its apparent concentration. Post-hybridisation washes are carried out first using
a low-stringency solution containing high salt levels, followed by sequential washes in lower salt solutions to increase stringency. This ensures that the DNA probe only stays bound to identical target sequences and reduces binding to homologous non-identical targets [Kelsell, 1996].

Membranes were placed in 20 ml of pre-hybridisation buffer at 40-42° C for 1-2 hours. This contained 50% de-ionised formamide, 5x Denhardt’s solution, 5x SSC, 10% dextran sulphate, 0.1 % SDS and 200 µl denatured salmon sperm DNA. (For Denhardt’s solution see Appendix III). Forty to fifty µl of labelled cDNA probe was added at 1-5 x 10^6 cpm/ml and incubated at the same temperature for a further 12-16 hours. Post-hybridisation washes were carried out as follows in volumes of 50-100 ml; two washes in 1x SSC/0.1% SDS at 40-42° C for 30 minutes, followed by a single wash in 0.25x SSC/0.1% SDS at 50° C for 15 minutes. Membranes were wrapped in Saran® film and exposed to autoradiography film with an intensifying screen at -80° C overnight.

### 3.9.3 Stripping of probe from membrane

Following development of the autoradiograph, the probe was removed from the membrane by placing it into boiling 0.1x SSC/0.25% SDS for 10-15 minutes. This was repeated as many times as necessary until a minimal radioactive signal was detected from the membrane, as measured with a Geiger counter. The membrane was then wrapped in Saran® film and stored at room temperature.
3.10 Solid phase RT-PCR using magnetic oligo(dT) beads

Dynabeads are magnetic beads coated with oligo(dT) sequences and can be used to capture poly(A) mRNA from total RNA preparations or directly from cell/tissue lysates. The oligo(dT) can also be used as a primer for reverse transcriptase and produce cDNA that remains covalently bound to the beads. This is known as solid phase RT-PCR. See figure 3.4.

Pellets of $1 \times 10^6$ cells from L-132 and SKLU-1 cell lines were used as templates for poly(A) mRNA extraction. Each cell pellet was mixed with 300 µl lysis buffer (100 mM Tris-HCl pH 8.0, 500 mM LiCl, 10 mM EDTA pH 8.0, 1% SDS, 5 mM DTT) and passed forcefully through a 21 gauge needle with a 2 ml syringe to shear the genomic DNA. The lysate was then mixed with 100 µl of Dynabeads, these had first been washed in lysis buffer, and left to anneal at room temperature for 5 minutes. The bead/mRNA complex was captured using a magnetic particle concentrator (MPC) and the supernatant removed. The complex was then washed once in wash buffer with SDS and once in wash buffer without SDS (10 mM Tris-HCl pH 8.0, 150 mM LiCl, 1 mM EDTA, +/- 0.1% SDS). The complex was then washed three times in cold RT reaction buffer. These washing steps remove any detergent and salts from the captured mRNA. The RT buffer was then removed via the MPC, leaving only the bead/mRNA complex, which was added directly to the RT-PCR mix. The reaction contained the following, 10 µl 5x RT buffer, 4 µl 0.1 M DTT, 2 µl 10 mM dNTP mix and 1 µl RNasin® in a final volume of 49 µl. This was incubated at 37 °C for 10 minutes before 1 µl M-MLV reverse transcriptase was added. The reaction was then left for a further 50 minutes at this temperature. The bead/mRNA/cDNA complex was captured from the mix using the MPC, resuspended in 50 µl Tris-HCl pH 8.0 and heated at 95 °C for 1 minute to melt the mRNA. The bead/cDNA complex was then captured using the MPC and resuspended in 50 µl Tris-HCl pH 8.0. This solution was used in the PCR step.

The captured bead/cDNA complex was washed in 50 µl PCR reaction buffer and 8.5 µl was used as the template for PCR. The reaction mix contained the following, 5 µl 10x PCR buffer, 1 µl 10 mM dNTP mix, primers (final concentration 1 pmol/µl), 0.5 µl Taq polymerase (5 U/µl) in a final volume of 50 µl.
Figure 3.4. Principles of solid phase RT-PCR using magnetic beads

Redrawn from Dynal® protocol handbook “Biomagnetic techniques in molecular biology”.

117
Chapter 3  Methods

3.11 RACE-PCR

RACE-PCR (rapid amplification of cDNA ends), also known as anchor-PCR, is a popular method of obtaining full-length cDNA sequences. It is a form of RT-PCR which amplifies sequences between a known region of the target mRNA and the unknown sequence at either the 5' or 3' end. The region of known sequence allows the design of a gene-specific primer (GSP) which is used in conjunction with a 5' or 3' anchor primer (adaptor) that has been ligated onto the target DNA strand. H8 5' RACE-PCR was achieved using the Marathon™ cDNA amplification protocol.

3.11.1 First & second-strand cDNA synthesis

A first-strand cDNA synthesis reaction was set up using 1µg poly A RNA, 10 µM cDNA synthesis primer, 5x first strand buffer, 10 mM dNTP mix, 100 U M-MLV enzyme in a final volume of 10 µl. The mix was incubated at 42° C for 1 hour before proceeding directly to second-strand synthesis. The entire first-strand reaction volume was added to 5x second-strand buffer, 10 mM dNTPs and 20x second-strand enzyme cocktail (DNA polymerase I, DNA ligase & RNase H) in a final volume of 80 µl. The reaction was incubated at 16° C for 90 minutes before adding 10 U T4 DNA polymerase and was then incubated for a further 45 minutes. The reaction was terminated by the addition of EDTA/glycogen and the cDNA was purified by phenol/chloroform extraction and precipitated with 4 M ammonium acetate and 95% ethanol. The sample was centrifuged at 13,500 rpm for 20 minutes at room temperature to pellet the cDNA, and this was then washed in 80% ethanol, air-dried and resuspended in 10 µl of H2O.

3.11.2 Adaptor ligation reaction

Five µl of the double-stranded cDNA solution was added to the ligation mix with 10 µM Marathon cDNA adaptor, 5x ligation buffer and 1U T4 DNA ligase in a final volume of 10 µl. This was then left at 16° C overnight. The ligase was heat inactivated and the adaptor-ligated cDNA was denatured and stored at -20° C until required for use.
3.11.3 5' RACE-PCR

Five μl of adaptor-ligated cDNA was used for PCR. The reaction also contained the following; 10 M adaptor primer, 10 μM 5'-GSP, 10x PCR buffer, 10 mM dNTPs, 1 U Taq DNA polymerase in a final volume of 50 μl. The reaction mix was overlaid with mineral oil and placed in a Perkin Elmer Cetus 480 Thermal Cycler programmed for 25 cycles of 94° C for 60 seconds, 60° C for 30 seconds and 68° C for 4 minutes. On completion of thermal cycling, 5 μl of the sample was analysed on a 1.2% agarose gel with appropriate size markers. The PCR product was then immediately ligated into the TA Cloning® vector and subcloned, as described above. Several white colonies were selected and incubated overnight in LB broth/ampicillin for use in DNA extraction and sequencing.
3.12 In situ hybridisation

The principle of in situ hybridisation is very similar to that of probe hybridisation using membrane-bound target RNA (or DNA). The only difference here is that the target RNA is still within the cells in a tissue section mounted on a glass slide. The section must first be treated to remove the paraffin wax in which it has been fixed. The exposed cells within the section must then be gently opened up by digestive enzymes so that the probe can enter and bind with its target. Various blocking agents are present to prevent non-specific binding and maximise the chances of successful hybridisation, similar to those described in section 3.9. Once the hybridisation reaction has occurred, the sections are then washed to remove any unbound or non-specifically bound probe, just as before.

This reaction does not necessarily require the use of a radiolabelled probe. Better resolution and faster signal development are achieved with non-radioactive probes, such as those labelled with digoxigenin or biotin. In this experiment the probe cocktail was labelled by the incorporation of a nucleotide which has been labelled with digoxigenin. Probe hybridisation is visualised by a chemical reaction catalysed by the enzyme alkaline phosphatase. The enzyme acts on the digoxigenin, in the presence of substrates molecules, to produce a coloured compound which can be seen within the cells under the light microscope. This method was developed by Pringle et al, 1987.

3.12.1 Oligonucleotide probe design

The H8 sense/coding strand was used as the template to generate anti-sense oligonucleotide sequences for potential use as probes using the G(C)G PRIMER package. The default parameters were altered to produce oligonucleotides 30-mers with a T\text{\textsubscript{melt}} of 80 °C and a minimum GC content of 20%. Potential probe sequences were selected with an average T\text{\textsubscript{melt}} of 65 °C, GC content of 40% and minimum 3' self-complementarity ratios.

Having satisfied all these criteria, potential sequences must then be checked for any homology against database sequences. Oligos with high levels of similarity to other sequences are unsuitable for use as they will cause non-specific hybridisation signals. G(C)G
FINDPATTERNS program was used to do this and the probe sequences were checked against the EMBL human database. This program has the advantage that many oligos can be checked simultaneously, rather than using BLAST or FASTA searches which only allow one oligo sequence to be checked per search. Selected oligonucleotide probe sequences were labelled with digoxigenin-11-dUTP as follows.

### 3.12.2 Oligonucleotide probe labelling

During this reaction, terminal deoxytransferase catalyses the addition of digoxigenin-labelled dUTP to the 3' end of the probe sequences.

One μg of each oligo was mixed together to give an oligo “cocktail”. One μg of this cocktail was added to the labelling reaction which also contained 4.0 μl 5x buffer [Boehringer], 4.0 μl 25 mM CoCl₂, 1.7 μl 1 mM digoxigenin-11-dUTP [Boehringer], 1.7 μl 5 mM dATP, 3.0 μl recombinant terminal deoxytransferase (15 U/μl) [Gibco] and dH₂O to a final volume of 20.0 μl. The reaction was mixed and left at 37 °C for 15 minutes before being stopped by the addition of 1 μl 0.5 M EDTA. The final volume was then made up to 50 μl with dH₂O.

Assuming 100% efficiency, there is now 1 pg labelled probe in 50 μl, i.e. 20 ng/μl. Most oligonucleotide ISH probes are used at ~200-500 ng/ml final concentration. The probe was then diluted to 500 ng/ml by adding 2.5 μl probe solution to 997.5 μl pre-hybridisation solution (see below for recipe) and stored at 4 °C until use. No separation of unlabelled probe is necessary.

### 3.12.3 Pre-treatment of slides and hybridisation

DEPC-treated solutions and RNase-free glassware were used at all stages here. Sections were kept clean and moist by soaking in DEPC.H₂O or being placed in a humidity box during or between stages, as appropriate.

Slides were dewaxed in xylene, 99% IMS and 95% IMS before being brought down to DEPC.H₂O. The slides were immersed in pre-warmed 2x SSC in an RNase-free container and incubated at 70 °C for 10 minutes, then washed in DEPC.H₂O for 5 minutes at room temperature. The sections were digested with 200 μl proteinase K (1-10 μg/ml) at 37 °C for 1
hour. The slides were then washed twice in DEPC.H₂O at 4 °C for 5 minutes. Following this, they were soaked in 0.4% paraformaldehyde/PBS at 4 °C for 20 minutes, before being washed once more in DEPC.H₂O at room temperature for 5 minutes. The sections are then incubated for 1 hour in pre-hybridisation solution at 37 °C. This solution contained the following: 0.6M NaCl, 1x PE (modified), 50% dextran sulphate, 150 µg/ml boiled salmon sperm DNA, 30% formamide and DEPC.H₂O. See Appendix III for PE (modified) recipe. The pre-hybridisation solution was replaced with an identical solution also containing labelled probe. The sections were covered with siliconized coverslips and incubated overnight at 37 °C.

3.12.4 Post-hybridisation washes and detection

The sections were washed in a solution containing 2x SSC/30% formamide and left at 37 °C for 10 minutes. This was performed twice. The sections were then immersed in an antibody blocking solution containing 3% BSA, 1% Triton-X-100 made up in 1x TBS and left at 37 °C for 5 minutes. The blocking solution was carefully removed from the sections. Antidigoxigenin antibody was diluted 1:600 in blocking solution and a volume of ~50-100 µl was added to cover each section. The sections were then left at room temperature for 30 minutes before being briefly washed twice in 1x TBS. The TBS was replaced with substrate buffer, containing 100 mM tris (pH 9.5), 5 mM MgCl₂ and 100 mM NaCl. The following reagents were added to 1 ml of the substrate buffer; 8 µl NBT, 8 µl BCIP and 1 µl levamisole. It should be noted here that these reagents are particularly toxic. NBT and BCIP act as substrates for the enzyme alkaline phosphatase and are necessary for the colorimetric reaction. Levamisole inhibits endogenous isoforms of this enzyme and so reduces background signal production. Approximately 100 µl of this buffer/substrate solution was added to cover each section. Non-siliconized coverslips were placed over the slides and they were then left in darkness at room temperature for 3 hours. The slides were examined after this time to make sure a colour change was occurring. They were then left for a further 12-16 hours.

To stop the reaction, the slides were gently washed in running tap water. They were counterstained with haematoxylin and mounted in an aqueous medium, such as Aquamount. It is very important that an aqueous mountant is used as the coloured precipitate that is produced by the ISH reaction is soluble in alcohol. The slides were then left to dry overnight and examined under a light microscope. A positive result is indicated by the presence of dark
brown/ black granules within the cells, while the background consists of paler brown areas. These colours will change from brown to blue over a period of ~1-2 weeks.

3.12.5 Controls for ISH

Several control slides should be included in every ISH reaction. A positive control tissue known to express the target sequence should be placed into the experiment. Another positive control is the use of a probe for common target sequences, such as mitochondrial or histone probes, or an oligo d(T) probe. Negative controls are equally important to ensure the results observed are real and not caused by background or other non-specific signals. A section of tissue known not to contain the target sequence is an obvious negative control. The use of test tissue sections digested with RNase to destroy the target is another common negative control. Sense-strand oligo probes should be identical to the target sequence rather than complementary to it and, thus, should not hybridise. Other negative controls are possible, such as using unlabelled anti-sense probe, using labelled vector only, using an irrelevant probe, and omitting key reaction components such as the probe or primary antibody.

In this experiment a mitochondrial probe was used as a positive control, while RNase treated sections and an absence of probe in the hybridisation solution were used as negative controls.
3.13 Computer software

Several computer programs were used during this project. Most were accessed through the molecular biology database (MOL) program package held on a UNIX based system called IRIX. The MOL menu contains several groups of programs, the largest of which is G(C)G. Each group has its own menu and on-line instructions for use. The programs used are listed below.

Table 3.2 Computer software programs used

<table>
<thead>
<tr>
<th>Program:</th>
<th>Source:</th>
<th>Use:</th>
</tr>
</thead>
<tbody>
<tr>
<td>(D)ATABASE/BLAST</td>
<td>IRIX/MOL</td>
<td>database search for sequence homology</td>
</tr>
<tr>
<td>(D)ATABASE/FASTA</td>
<td>IRIX/MOL</td>
<td>database search for sequence homology</td>
</tr>
<tr>
<td>G(C)G/GELMERGE</td>
<td>IRIX/MOL</td>
<td>alignment of multiple sequences to look for areas overlap</td>
</tr>
<tr>
<td>G(C)G/GELASSEMBLE</td>
<td>IRIX/MOL</td>
<td>editor program to assemble consensus sequence</td>
</tr>
<tr>
<td>G(C)G/PEPDATA</td>
<td>IRIX/MOL</td>
<td>translates sequence in all possible reading frames</td>
</tr>
<tr>
<td>G(C)G/MAP</td>
<td>IRIX/MOL</td>
<td>restriction digest map</td>
</tr>
<tr>
<td>(P)RIMER</td>
<td>IRIX/MOL</td>
<td>primer design</td>
</tr>
<tr>
<td>Seq.Ed v1.0</td>
<td>Mac</td>
<td>sequence analysis for automated sequencing</td>
</tr>
<tr>
<td>UV/VIS</td>
<td>PE</td>
<td>optical densitometry readings for DNA/RNA</td>
</tr>
<tr>
<td>ImageQuant</td>
<td>Mol Dyn</td>
<td>scanning &amp; image analysis for autoradiographs</td>
</tr>
<tr>
<td>RepeatMasker2</td>
<td>Internet*</td>
<td>identifies repeat regions within a sequence</td>
</tr>
</tbody>
</table>

Key:  Mac = Apple MacIntosh program  
PE = Perkin Elmer program  
Mol Dyn = Molecular Dynamics program  
* Website = http://ftp.genome.washington.edu/cgi-bin/RepeatMasker
The following Internet web sites were also used as general sources of information and, thus, are not specifically mentioned in the text.

*Lung cancer & smoking*
http://ourworld.compuserve.com/homepages/LungCancer

*Pulmonary pathology index*
http://www-medlib.med.utah.edu/WebPath/LUNGTML/LUNGIDX.html

*Medicine on-line lung cancer information center*
http://indy.kbt.com/wn/leukaemia/webmate/leukaemia/form/leukaemia/lunginfo

*National Center for Biotechnology Information*

*Directory of molecular biology resources*
http://www.mc.vanderbilt.edu/resources/interests/molecular/directories.html
Chapter 4 - Adaptation and development of the DDRT-PCR technique for use with human lung RNA
4.1 Strategy for development of DDRT-PCR method for use with human lung RNA

The first, and most important, task of the study was to adapt and develop the DDRT-PCR methodology for use with human lung RNA. This was achieved in 4 stages;

i] Using control RNA samples & manufactured DDRT-PCR kit,
ii] Using lung cell line RNA samples & manufactured DDRT-PCR kit,
iii] Using lung cell line RNA samples & “home-made” DDRT-PCR reagents,
iv] Using human tissue RNA samples & “home-made” DDRT-PCR reagents.

The first stage was a relatively straightforward process of setting up and becoming familiar with the methodology. Control RNA extracted from rat liver and mouse lung was used in addition to RNA from mouse fibroblasts for initial experiments. The latter was provided with the DDRT-PCR kit (RNA map kit A, Biogene). These controls were used to determine, firstly, if the method would work and, secondly, to verify that RNA sources from different animals and different tissues would produce distinct banding patterns. Clear banding patterns were produced from each control RNA template, and numerous banding differences were apparent between the three samples. These findings are illustrated in figure 4.1, in which each control RNA sample was amplified with all four anchor primers in combination with arbitrary primer AP1. Some variation in the number of bands attained from the templates was observed but the number of bands per lane was within the range expected, i.e. 75-150. Also, defined banding patterns from each template were obvious. (See also later figures in Chapters 5 and 6.)

The next stage of the methodology was to obtain banding patterns with human lung cell line RNA samples and kit reagents. Preliminary experiments were set up using RNA extracted from human lung epithelial cell line L-132 representing the normal specimen, and from human lung adenocarcinoma cell line SKLU-1 representing the tumour specimen. RNA extractions were performed using RNAzol™ B reagent. All RNA was tested for its ability to be amplified by undergoing RT-PCR using K-ras primers before any DDRT-PCR was attempted. Ras primers were chosen for this purpose because there is a high degree of evolutionary conservation in this family of genes, implying an essential role for ras proteins in normal cell physiology [Leon et al, 1987, Hesketh, 1994]. See Chapter 3, section 3.2.4 for methodology.
Figure 4.1

An autoradiograph showing segments of the differential banding patterns obtained from RNA samples extracted from mouse lung tissue, rat liver tissue and mouse fibroblast cells. The differences in banding patterns produced from each template with the AP1/ T12MN primer combinations can clearly be seen.
Figure 4.1 DDRT-PCR using control RNA samples with primer API

<table>
<thead>
<tr>
<th></th>
<th>$T_{12}MA$</th>
<th>$T_{12}MC$</th>
<th>$T_{12}MG$</th>
<th>$T_{12}MT$</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

M = Mouse lung RNA
R = Rat liver RNA
F = Mouse fibroblast RNA
and Appendix II for primer details. Any RNA sample whose cDNA failed to be amplified with these primers would be unlikely to contain cDNAs for less common messages and, thus, be of little use in DDRT-PCR. Mouse fibroblast RNA remained in use as a positive control for the DDRT-PCR. Again, no major problems were encountered, save for minor alterations in the starting amounts of RNA used in the RT-PCR step of the DDRT-PCR.

It was then decided to try to substitute “home-made” reagents for kit reagents to see if the process could be made more cost-effective. Reagents were replaced one by one, being constantly compared to reactions containing their kit counterparts to ensure no reduction in specificity or in sensitivity occurred. It was discovered that the enzymes used in DNase treatment of the RNA and in the cDNA synthesis steps could be removed without any ill effect, as long as their replacements were from Perkin Elmer, Sigma or Gibco. Arbitrary and anchor primers were synthesised “in house” and, providing they were purified and aliquoted with care, no problems were observed during their use.

The final phase of the experimental set-up was to use matched human lung adenocarcinoma tissues from which the control and test RNA samples would be produced. Five matched cases were used to prepare total RNA. The quality of the tissue RNAzol™ B preparations was observed to be inferior to that produced from the cell line samples, particularly for the normal lung tissue. Nevertheless, all the tissue RNA samples were placed in control RT-PCR amplifications with K-ras primers. As these amplifications were successful, the RNA samples were then used in the DDRT-PCR with arbitrary primer #1.
4.2 Results and trouble-shooting experiments for DDRT-PCR using human lung RNA

It was during the final phase of the experimental development that the first significant difficulties of the study occurred, and three main problems were encountered. Firstly, lanes were missing from the display at random, with duplicates behaving in an erratic manner. Secondly, the normal lung RNA samples were producing very much weaker displays than the tumour RNA samples. Finally, the gradual and persistent degradation of lung tissue RNA samples was noted. A series of experiments were designed to address the problems encountered with the human tissue RNA samples. All factors pertaining to each difficulty were examined in turn and in a logical order. See figure 4.2.

4.2.1 Investigation of “ghost” lanes in DDRT-PCR gels

The first major anomaly that was noted to occur over several months was the seemingly random loss of signal in some lanes of the PCR gel. The problem was most obvious when one of a pair of duplicate samples produced a banding pattern while its partner produced a lane that was completely blank, figure 4.3. The reasons for this remained unclear but possibilities included pipetting errors, random contamination/ degradation of the template or of one of the reaction components, and malfunction of the PCR thermal cycler. All reagents were re-made and re-aliquoted in a class II cabinet using sterile technique to avoid contamination by any inhibitory factors. Fresh enzyme preparations, nucleotide solutions and isotope were used at all times. A series of technical problems involving the PCR heating blocks did occur which may have had an effect on subsequent results.

The ambient temperature in the laboratory was in excess of 35° C for several weeks during the unusually hot summer weather of 1995, the time that this work was undertaken, which caused the PCR thermal cyclers sometimes to overheat. The PCR machines used in the Unit were all of the “hot-block” design, controlled by oil-filled heat exchangers which are cooled by a fan. This type of machine is not designed to cope with extremes of temperature and is constructed to function optimally in ambient temperatures not exceeding 28° C. The thermal cyclers were eventually moved into an air-conditioned room but, unfortunately, the air-
Figure 4.2. Summary of action for DDRT-PCR trouble-shooting experiments.

Problems

Ghost lanes ← RNA degradation

Unequal DDRT-PCR signal

RNA samples → Degrading → Intrinsic lung factors? (RNase / collection)

→ RNA extraction method? (method 1/2)

→ Contamination of reagent(s)?

→ RNA storage methods?

→ Not degrading → Concentrations insufficient?

→ DDRT-PCR problem?

DDRT-PCR

→ amount starting RNA in RT step to ↑[cDNA]?

→ amount cDNA in PCR step to ↑[PCR product]?

→ labelling of PCR product? (³⁵S/³²P...)

→ PCR efficiency optimal? (Mg²⁺ / dNTPs...)

→ PCR components at fault? (Hot block / tubes...)
Figure 4.3

Selected lanes from different DDRT-PCR autoradiographs demonstrating the failure of one in a pair of duplicate reactions. This phenomenon was called “ghosting” and the lanes in which it occurred were referred to as “ghost” lanes. The precise reason for this occurrence was not determined.
Figure 4.3 “Ghost” lanes on DDRT-PCR gel

In each case, one of the duplicates has failed to amplify (x) while the other has produced a more usual banding pattern (√)

N= normal cDNA sample
Ca= tumour cDNA sample
conditioner units also overheated! The problem remained apparent until the weather returned to a more traditional British summer climate!

The positions of individual reaction tubes in the PCR blocks were monitored to determine if reactions at certain locations were failing. No consistent patterns were observed. These “ghost” lanes remained a significant impediment to the experiments. It has been suggested that minor alterations in reaction efficiency may be caused by variations in the reaction tubes used [Chen, 1994]. It was found that differences between tubes from different manufacturers and even between different batches from the same manufacturer could result in one or a few missing reactions, with one reaction of a duplicate sample working while the other failed. The same source of PCR tubes was used for most of these differential display reactions, although the individual batches did vary. Towards the end of this phase of experiments, thin-walled PCR tubes were used from a different supplier but it is difficult to say whether this had any effect on reaction efficiency. It is of interest to note that other workers in the Unit using DDRT-PCR also experienced similar problems [personal communications].

4.2.2 Attempts to increase differential display band signal

The signal on the DDRT-PCR gels from lanes containing cDNA made from normal RNAs was much weaker than that from the tumour samples. This made banding pattern comparisons very difficult. A series of investigations was undertaken to correct this problem.

4.2.2.1 Improving RNA extraction & quantification

The yield of RNA from the normal tissue was consistently lower than that obtained from a similar sized portion of tumour tissue. Spectrometer readings of RNA samples following DNase treatment gave concentrations of 250-1400 μg/ml for the controls and 500-1800 μg/ml for the tumour samples, see table 4.1. Quantification by gel electrophoresis generally revealed this to be an overestimate in the amount of RNA present. It was, therefore, conceivable that much smaller amounts of RNA had been introduced in the RT step than had been appreciated. It was possible that phenol contamination from the DNase step was responsible for these falsely high readings. As a result, future spectrometer readings were used only as a guide to RNA quality and quantification was achieved by comparison of loading
volumes via agarose gel electrophoresis. Gel electrophoresis also demonstrated a deterioration in the quality of the RNA samples, as ribosomal RNA 18 and 28 S bands were no longer visible in some cases.

Table 4.1. Human tissue RNA sample OD readings

<table>
<thead>
<tr>
<th>Sample</th>
<th>Method</th>
<th>A\textsubscript{260/280}</th>
<th>µg/ml</th>
<th>DNase</th>
<th>A\textsubscript{260/280}</th>
<th>µg/ml</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1- Ca</td>
<td>RNAzol</td>
<td>1.9</td>
<td>8100</td>
<td>YES</td>
<td>1.8</td>
<td>1800</td>
</tr>
<tr>
<td>#1- N</td>
<td>RNAzol</td>
<td>1.9</td>
<td>2700</td>
<td>YES</td>
<td>1.8</td>
<td>1400</td>
</tr>
<tr>
<td>#2- Ca</td>
<td>RNAzol</td>
<td>2.0</td>
<td>6000</td>
<td>YES</td>
<td>2.0</td>
<td>500</td>
</tr>
<tr>
<td>#2- N</td>
<td>RNAzol</td>
<td>2.0</td>
<td>1800</td>
<td>YES</td>
<td>1.8</td>
<td>250</td>
</tr>
<tr>
<td>#3- Ca</td>
<td>RNAzol</td>
<td>1.2</td>
<td>720</td>
<td>N/D</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>#3- N</td>
<td>RNAzol</td>
<td>1.6</td>
<td>1900</td>
<td>N/D</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>#4- Ca</td>
<td>RNAzol</td>
<td>1.1</td>
<td>700</td>
<td>N/D</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>#4- N</td>
<td>RNAzol</td>
<td>1.8</td>
<td>210</td>
<td>N/D</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>#5- Ca</td>
<td>RNAzol</td>
<td>1.8</td>
<td>490</td>
<td>N/D</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>#5- N</td>
<td>RNAzol</td>
<td>1.7</td>
<td>1000</td>
<td>N/D</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Key: Ca = tumour tissue RNA  
N = normal tissue RNA  
N/D = not done (see text below)  
RNAzol = RNAzol™ B extraction method  
A\textsubscript{260/280} = purity measurement ("clean" RNA should have A\textsubscript{260/280} of 1.8-2.0)  

[Please note figures are based on extraction from ~50 mg tissue in final volumes of 20-50 µl]

All RNA samples were re-made according to the RNAzol™ B protocol. As stocks of tissue were limited, smaller starting quantities had to be used for RNA extraction and yields were, therefore, also reduced. In order to preserve as much of the RNA sample as possible it was decided to minimise the number of manipulations imposed upon each sample. Thus, the DNase purification step was omitted (cases 3-5) as a significant proportion of the RNA sample was lost during this particular procedure. However, the signal strength of the display pattern did not show any discernible improvement. Attempts were then made to try to increase the
amount of cDNA produced in the RT-step by using twice the quantity of RNA for the cDNA synthesis. See figure 4.4. Similar efforts were instigated to improve the production of DD-PCR product by adding double the amount of cDNA to the PCR step. In many lanes a faint pattern was present but it was not of sufficient intensity to allow reliable comparisons to be made between the samples. See figure 4.5. If PCR products were present, albeit dimly, was there a problem with the labelling?

4.2.2.2 Improving labelling efficiency for the DD-PCR products

Efforts were then focused on improving isotope incorporation and increasing the labelling intensity of the DDRT-PCR products. Nucleotide mixes with various decreasing amounts of unlabelled dATP (12.5-2.5 μM) were prepared and used in the PCR step to try to promote increased labelling by α\(^{35}\)S-dATP. Additional amounts of Taq DNA polymerase were also added to the reactions after 20 cycles to try to increase product formation. Neither of these strategies made any discernible difference to the banding pattern signal. The isotope was then replaced with α\(^{32}\)P-dCTP in the hope that its greater activity would provide a stronger signal for the display. Although both \(^{35}\)S and \(^{32}\)P produce the same type of emissions (β), the maximum energy from \(^{32}\)P emissions is much higher than that from \(^{35}\)S, 1.71 MeV for \(^{32}\)P against 0.167 MeV for \(^{35}\)S [Mundy et al, 1994]. This means that DNA labelled with \(^{32}\)P is “hotter” than that labelled with \(^{35}\)S, i.e. produces a stronger signal. The disadvantage of using \(^{32}\)P is that it has a lower achievable resolution than \(^{35}\)S, 200-500 μm against 100-200 μm, respectively [Mundy et al, 1994], so the signal from \(^{32}\)P will be more diffuse than that from \(^{35}\)S. When α\(^{32}\)P-dCTP was used in the reaction although the strength of some bands was increased, the overall pattern obtained remained poor and band definition was reduced, giving a “fuzzy” appearance to the gel patterns. See figure 4.6.

4.2.2.3 Increasing PCR efficiency

As the efficiency of the PCR seemed to have diminished since the establishment of the method, it was decided to carry out a series of Mg\(^{2+}\) and dNTP titrations to confirm optimal PCR conditions and maximise enzyme performance. The final concentration of dNTPs in the original protocol was 2.5 μM. This relatively low amount was selected by the method’s
Figure 4.4

Autoradiographs from two cases where greater amounts of RNA were placed into the cDNA synthesis step of the reverse transcription reaction of DDRT-PCR.

This was done in an attempt to produce higher concentrations of cDNA in the hope that it would improve the yield of PCR products obtained. This in turn was intended to produce a better banding signal when these products were run out on a denaturing gel. It was hoped especially that this would enhance the signals from the normal (N) lanes, as they were all persistently very faint.

However, it can be seen that this strategy has made negligible difference to the resulting signal.
Figure 4.4 DDRT-PCR gel demonstrating a lack of improvement in signal from the normal RNA lanes when using increased RNA template in the first-strand synthesis RT step.

<table>
<thead>
<tr>
<th></th>
<th>T_{12}MA</th>
<th>T_{12}MC</th>
<th>T_{12}MG</th>
<th>T_{12}MT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td>Ca N</td>
<td>Ca N</td>
<td>Ca N</td>
<td>Ca N</td>
</tr>
<tr>
<td>Case 4</td>
<td>Ca N</td>
<td>Ca N</td>
<td>Ca N</td>
<td>Ca N</td>
</tr>
</tbody>
</table>

Key: N = normal cDNA  Ca = tumour cDNA  
T_{12}MA/C/G/T = anchor primers  AP1/5 = arbitrary primers
Figure 4.5

Autoradiographs from one case where greater amounts of cDNA were placed into the PCR step of DDRT-PCR.

The strategy was similar to that used in figure 4.4, that was to try to improve the yield of PCR products obtained and produce stronger banding patterns, particularly in the weak normal (N) lanes.

However, it can be seen that this strategy has made negligible difference to the resulting signal.
Figure 4.5  DDRT-PCR gel demonstrating a lack of improvement in signal from normal RNA lanes when using increased cDNA template in PCR step.

\[ N = \text{normal tissue cDNA} \]
\[ Ca = \text{tumour tissue cDNA} \]

The faint signal in the normal cDNA lanes has not increased despite the use of greater cDNA template in the PCR step.
**Figure 4.6**

Autoradiographs from two experiments comparing signal strength of PCR products labelled with different isotopes. $^{32}$P labelled nucleotide was incorporated into the PCR products and compared to products containing $^{35}$S labelled nucleotide.

As in figures 4.4 and 4.5, the intention was to increase the banding pattern signal, particularly for the normal (N) lanes. The bands labelled with $^{32}$P were much stronger than those labelled with $^{35}$S, but they lacked the clarity and resolution present in the latter. Also, there was no improvement in the signal from the lanes derived from the normal cDNA template.
Figure 4.6  Selected lanes from a DDRT-PCR gel comparing the signal obtained from the use of $^{32}$P and $^{35}$S labelled nucleotide.

$^{32}$P-dCTP labelling

(T$_{12}$MA x AP5)

$^{35}$S-dATP labelling

(T$_{12}$MA x AP5)

Note the diminished definition and resolution of bands in the $^{32}$P lanes compared with those in the $^{35}$S lanes. Arrows indicate equivalent band positions.

Key:  
N = normal cDNA  
Ca = tumour cDNA  
T$_{12}$MA = anchor primer  
AP5 = arbitrary primer
originators to minimise background banding in the cDNAs used [Liang and Pardee, 1992]. As the cDNAs used here were different and the occurrence of excessive background banding had definitely not been a problem, it was decided to increase dNTP concentrations. The dNTPs bind Mg$^{2+}$ ions, which lowers the amount of free ions available for use as a co-factor for the enzyme Taq DNA polymerase [Oste, 1989]. If more dNTP is used, then more Mg$^{2+}$ must also be added. Mixes of dNTP at final concentrations of 2.5, 10 and 25 μM were used in the PCR step, with enzyme buffer containing Mg$^{2+}$ at final concentrations of 1.5, 2.5 and 4.0 μM. Results demonstrated optimal banding (i.e. maximal number with minimum background) at a final Mg$^{2+}$ concentration of 1.5 μM with a final dNTP concentration of either 10 or 25 μM. See figure 4.7. Interestingly, these final dNTP concentrations were 4-10 times greater those originally used by Liang and Pardee and one would also have expected a higher Mg$^{2+}$ requirement.

Experiments were repeated using fresh RNA samples and the modified PCR conditions. Unfortunately, no improvement in the quality of results was observed. The RNA samples were re-checked for quality and for an ability to be amplified. All were found to be in a state of degradation and incapable of amplification, even by K-ras primers. This provoked the question of whether it was possible to extract good quality RNA from any of the lung tissue samples collected? Experiments were then designed to determine whether the problem was intrinsic to these samples, or was due to a more general problem with lung tissue per se. Also, to assess equipment and reagents for contamination by RNase.

4.2.3 Investigation of persistent lung tissue RNA degradation

As noted above, once lung tissue replaced the lung cell lines as the source of RNA, the stability of the RNA was noted to deteriorate. The cause of this apparently inevitable degradation was not immediately clear. Was it due to intrinsic factors within the tissue samples collected? Was it due to a more general problem with lung tissue? Or was it due to extrinsic factors in the mode of preparation? A series of investigations was set up to answer each question in turn.
In an attempt to establish optimum conditions for DDRT-PCR, the reaction was carried out under various Mg\(^{2+}\) and dNTP concentrations. The amount of dNTPs in the reaction is very important, as too little will preclude amplification of sufficient product to be detected and too much will result in the appearance of non-specific bands. The Mg\(^{2+}\) ions act as co-factors for the enzyme *Taq* polymerase and are crucial for its successful functioning. The presence of dNTPs reduces the amount of free ions available for the enzyme, thus as dNTP concentration rises so must that of Mg\(^{2+}\).

Standard DDRT-PCR conditions used dNTPs at a final concentration of 2.5 \(\mu\)M and Mg\(^{2+}\) at 1.5 mM. Additional final concentrations of dNTP at 10 \(\mu\)M and 25 \(\mu\)M and of Mg\(^{2+}\) at 2.5 mM and 4.0 mM were also used included. Reactions were set up using each combination of dNTP and Mg\(^{2+}\) with template RNA from both normal (L, L-132) and tumour (S, SKLU-1) cell lines.

Optimal banding signal, maximum band number with minimal background, was obtained with the standard Mg\(^{2+}\) of 1.5 mM final concentration and dNTPs at a final concentration of 10 \(\mu\)M or 25 \(\mu\)M. Improvements were seen in both normal and tumour samples.
Figure 4.7 Mg$^{2+}$ / dNTP titration results.

Lanes 1+2 dNTPs = 2.5 µM (final concentration)
Lanes 3+4 dNTPs = 10 µM (final concentration)
Lanes 5+6 dNTPs = 25 µM (final concentration)
4.2.3.1 Comparison of RNA extraction methods

To preserve test-case tissue stocks RNA was prepared from one of the matched non-adenocarcinoma lung tumours collected, a squamous cell carcinoma. The rationale was as follows; if this gave the same results as the adenocarcinoma tissue it would imply the problem was general to lung, and if the results were different it would imply the problem was with the adenocarcinoma specimens. Two RNA extraction protocols, RNAzol™ B and acid GITC methods [Chomczynski and Sacchi, 1987], were also used to determine if the problem was with one of the reagents. When the samples were compared by agarose gel electrophoresis, the RNA prepared according to the RNAzol™ B protocol was revealed to be of extremely poor quality. It showed an absence of both 18 and 28S ribosomal RNA bands and lack of reliable amplification with K-ras primers. See figure 4.8. Samples of RNA prepared according to the acid-GITC protocol were vastly superior, however, demonstrating both 18 & 28S bands and the ability to amplify with RT-PCR and DD-PCR. These results demonstrated that it was possible to extract good quality RNA from lung tissue! Using the acid-GITC protocol, RNA was extracted from the test-case tissues once more. Case #3 failed to amplify with K-ras primers and was discarded, figure 4.9. The presence of a single, clean PCR product signal of ~ 290 bp from the amplification with K-ras primers also demonstrated the lack of contaminating genomic DNA in these RNA samples. The remaining 4 case RNA samples were used in DDRT-PCR with somewhat mixed results. Again, the tumour samples provided a good differential display pattern but the normal samples did not. The reason for the lack of ability of the normal tissue RNA to be amplified by DDRT-PCR remained unclear. Macroscopically, the control tissue was more fibrous, heavily blood-stained and contained more carbon pigment than the tumour tissue. It is possible, although unlikely, that one of these constituents was contributing to an inhibitory effect on the PCR. Any such contaminants should, however, have been removed during RNA extraction and purification. Although a dilution series would have been the logical investigation of any putative contaminant, there would have been insufficient cDNA remaining to amplify by DDRT-PCR and for this reason it was not pursued.
Agarose gel electrophoresis comparing the RNA from matched human squamous cell lung carcinoma tissue obtained using two different extraction techniques; the conventional acid GITC protocol and a single-step extraction using an “all in one” commercial reagent, called RNAzol™ B. [Approximately 5 μg RNA loaded per lane]

The upper gel (a) demonstrates the difference in the resulting RNA samples. Those obtained with acid GITC appear superior, with their ribosomal RNA 18S and 28S bands clearly visible. The samples from the RNAzol™ B extraction appear to be partially degraded and no ribosomal bands are obvious.

The lower gel (b) shows the fate of these samples when used in RT-PCR with K-ras primers. PCR product is clearly present in both of the acid GITC template samples. PCR product can only be seen in the tumour template sample extracted with RNAzol™ B.
Figure 4.8 Comparison of RNAzol™ B and acid GITC RNA extraction protocols.

a) RNA check gel

<table>
<thead>
<tr>
<th>Tumour RNA</th>
<th>Normal tissue RNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>R</td>
</tr>
<tr>
<td>A</td>
<td>R</td>
</tr>
</tbody>
</table>

5 μg total RNA, extracted from matched human squamous cell lung carcinoma tissue, was loaded in each lane. A clear difference in the quality of the resultant RNA can be seen. The RNAzol™ B samples are markedly inferior in quality (a) and in response to amplification by K-ras primers (b).

Key:  
A = Acid GITC extraction  
R = RNAzol™ B extraction  
Co = RT-PCR positive control
Figure 4.9

Agarose gel electrophoresis showing PCR products obtained from the amplification of cDNA derived from matched human adenocarcinoma cases 1-5.

(a) The RNA samples were placed in a reverse transcription reaction primed with anchor primers $T_{12}$MA/ C or G. The resulting cDNAs were then amplified with $K$-ras primers. This was used as a control to assess the quality of the RNA/ cDNA before it was used in DDRT-PCR.

(b) PCR products can be seen in those lanes corresponding to cases 1, 2, 4 and 5. In all cases, a less intense PCR product band was obtained from amplification of the normal cDNA than from the tumour cDNA. Case 3 failed to amplify, although a faint band is just visible in the normal cDNA/ $T_{12}$MC lane.
Figure 4.9 RT-PCR with K-ras primers using RNA from cases 1-5.

Case #1

<table>
<thead>
<tr>
<th>Case #1</th>
<th>Tumour cDNA</th>
<th>Normal tissue cDNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>cDNA</td>
<td>LS ACG Co</td>
<td>LS ACG Co</td>
</tr>
</tbody>
</table>

Cases #2 - 5

<table>
<thead>
<tr>
<th>Cases #2 - 5</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>LS ACG Co LS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Key:
- A/C/G = anchor primer T₁₂MA/C/G (used in first strand cDNA synthesis)
- Co = RT-PCR positive control
- LS = length standard (100bp ladder)
4.2.3.2 Comparison of RNA extraction from other tissues

It looked increasingly as though human lung tumour tissue RNA was not going to be suitable for use in the DDRT-PCR, so would animal lung tissue be any better? Samples of normal and adenoma lung tissue from urethane-treated BALB-C/AJ F1 mice were used for RNA preparation. Although the samples could be amplified by K-ras RT-PCR, results from DDRT-PCR were disappointing. Smearing in all lanes was observed, with a complete absence of bands. RNA was then extracted from samples of non-neoplastic human lung tissue, including cases of chronic obstructive airways disease, bronchiectasis, non-specific lung inflammation and a pulmonary cyst. The resultant RNAs failed even to amplify with K-ras primers and were immediately discarded. Finally, RNA was prepared from matched non-pulmonary adenocarcinoma tissues, using samples from pancreas, stomach, colon, ovary and liver. Although seemingly good quality RNA was obtained which amplified with K-ras, results from DDRT-PCR were again poor with smearing in lanes indicative of RNA degradation. See figure 4.10. The RNA samples made from the original test cases were re-checked and also found to be degrading. Other RNA samples were examined from both pulmonary and non-pulmonary origins, tumour and non-tumour tissues. In total, 32 RNA samples were checked and all were in varying state of degradation, the lung tissue samples being the worst. This was inspite of the samples being stored under different conditions (under liquid nitrogen or at -80 °C, in DEPC.H2O or ethanol, concentrated or dilute, in the presence or absence of RNase inhibitors). See figure 4.11.

4.2.3.3 Investigation of RNase contamination of reagents

The only other possibility which would explain this observation was RNase contamination of one or more of the reagents or pieces of equipment used during RNA preparation. If this proved to be correct, it was indeed unfortunate as every conceivable precaution had been undertaken at all stages of RNA preparation. (See Chapter 3, section 3.2 and 3.3.) There was no way of knowing where any contamination may have occurred, all that remained was to remake all reagents and clean all equipment. Fortunately, two new anti-RNase products had just become commercially available; RNaseZAP™ and RNaseALERT™ (Ambion, UK). The former is an alkaline solution used for cleaning equipment, work surfaces and tubes, while the latter is a dip-stick test for RNase contamination of reagents. All
Figure 4.10

(a) Agarose gel electrophoresis showing PCR products obtained from the amplification of cDNA derived from matched human non-pulmonary adenocarcinomas. This was done to assess the quality of the human non-lung tissue RNA samples. As in figure 4.9, the RNA samples were placed in a reverse transcription reaction primed with anchor primers $T_{12}MA/C$ or $G$, and the resulting cDNAs were amplified with $K$-$ras$ primers. Successful amplification was observed in each lane, implying the samples were suitable for use with DDRT-PCR.

(b) Autoradiograph of DDRT-PCR gel showing the resulting banding patterns obtained from differential display using the cDNAs from (a) with arbitrary primer AP$_1$.

No bands were visible in any lane, with the majority showing background signal only.
Figure 4.10 RT-PCR and DDRT-PCR results using RNA extracted from matched non-pulmonary human tissue samples.

a) RT-PCR with K-ras primers

<table>
<thead>
<tr>
<th>Pancreatic adenocarcinoma</th>
<th>Colonic adenocarcinoma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Co N Ca</td>
<td>A C G A C G</td>
</tr>
</tbody>
</table>

b) DDRT-PCR with arbitrary primer API

| Co A C G | Pancreatic adenocarcinoma A C G NCa NCa NCa | Colonic adenocarcinoma A C G NCa NCa NCa |

Key:  
N = normal tissue  
Ca = tumour tissue  
Co = positive control  
A/C/G = anchor primer T_{12}MA/C/G
Figure 4.11

Agarose gel electrophoresis showing the degradation of numerous RNA samples extracted from human tissues.

(a) Shows RNA samples extracted from the matched human adenocarcinoma cases 1-5, as well as RNA taken from some non-neoplastic lung pathologies.

(b) Shows RNA samples extracted from various sources and stored under different conditions.

All samples were in a state of degradation, regardless of origin, organ or storage.

[Please note that although the RNA ladders appear to be quite faint on this photograph, in reality the ladder markings were visible.]
Figure 4.11 Degradation in lung tissue RNA samples

a) Samples stored at -80 °C in DEPC.H₂O

<table>
<thead>
<tr>
<th>Case</th>
<th>#1</th>
<th>#2</th>
<th>#3</th>
<th>#4</th>
<th>#5</th>
</tr>
</thead>
<tbody>
<tr>
<td>LS</td>
<td>N</td>
<td>N</td>
<td>CaCa</td>
<td>N</td>
<td>Ca</td>
</tr>
</tbody>
</table>

Non-neoplastic lung lesions

b) Samples stored at higher concentration under liquid nitrogen or at -80 °C under 100% ethanol

<table>
<thead>
<tr>
<th>Liquid nitrogen samples</th>
<th>(-80 °C + Ethanol) samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case #3 #4 #5</td>
<td>#2 #3 #4 #5 / Mouse lung / Cell lines</td>
</tr>
<tr>
<td>LS N Ca N Ca N Ca</td>
<td>Ca Ca Ca Ca N Ca N Ca Adj N Ca LS</td>
</tr>
</tbody>
</table>

Key: N = normal tissue RNA (cell line L-132)
Ca = tumour tissue RNA (cell line SKLU-1)
Adj = normal tissue adjacent to tumour area
LS = RNA ladder
equipment was thoroughly cleaned and all solutions were tested. None was found to contain RNase, according to the dip-stick test. A final RNA extraction from lung tissue was performed using freshly prepared RNase-free solutions and equipment. Unfortunately, it degraded just as before.

4.3 Discussion

Lung tissue contains an admixture of cell types. Inflammatory cells, such as neutrophils and eosinophils, may often be present, particularly if there is a background of asthma, bronchitis, pneumonitis or fibrosis [Momex et al, 1994, Laskins and Pendino, 1995]. Some of these conditions, such as bronchitis, are associated with cigarette smoking and thus may be present in a lung cancer patient. Lung tumours may also provoke an inflammatory response, particularly if they result in a pneumonia or in abscess formation. Any inflammatory changes will cause an influx of inflammatory cells and the subsequent release of nucleases, including RNase. It is generally accepted that lung tissue contains high levels of RNase enzymes [Morita et al, 1986]. These nucleases may cause degradation of RNA during and after RNA isolation [Evans and Kamdar, 1990]. It is likely that residual nucleases were present in varying amounts in the RNA samples extracted from the human tissue cases and that they were responsible for the ensuing degradation observed. This would also explain why the initial RNA preparations from the human tissue samples appeared intact and performed well in RT-PCR with K-ras primers. The K-ras PCR product was only 300 bp long and it is possible that a partially degraded RNA sample could still contain sufficient unbroken target fragments to undergo successful amplification. In retrospect, perhaps a choice of primers that amplified a larger target, such as β-actin, would have been more useful as a guide to RNA quality. However, while good laboratory practices and precautions against RNase activity are of course important, if RNA is going to degrade it will degrade no matter what one does to try and prevent it! Recently, a modification to the RNAzol B protocol that claims to remove residual RNase from extracted RNA has been described [Kodavanti et al, 1996]. The modification involves an additional extraction with RNAzol reagent and chloroform which appears to remove inflammatory proteins that are resistant to conventional RNA purification procedures [Gleich et al. 1986].
Chapter 4  Adaptation and development of the DDRT-PCR technique for use with human lung RNA

It is possible that, even though the lung tissue samples were collected and frozen within thirty minutes of removal from the patient, RNases had already been released and damaged the RNA. Also, the test case tissues were collected towards the periphery of the lung, as this is where adenocarcinomas occur. Any ischaemic damage caused during surgery by blood vessel ligation would have the most immediate effect in this region as it is most distal to the pulmonary vessels, resulting in release of degrading enzymes. Despite attempts to keep the time required to get the surgical specimens frozen in liquid nitrogen to a minimum, it would still have taken the surgical team at least 15-30 minutes to mobilise the lung sufficiently to allow its removal from the patient once the pulmonary vessels had been disconnected, providing ample time for the initiation of RNA degradation. During examination of the gross specimens, many of the adenocarcinoma samples were noted to be friable and partially necrotic. It was likely, therefore, that RNA damage was present within these areas. The same reasoning could also be applied to the non-neoplastic lung RNA samples. The failure of the non-pulmonary adenocarcinoma RNAs to be differentially displayed is not as clear. Some of these tissue samples were several years old, thus the exact method of collection and conditions of storage cannot be verified. It may have been that when the tissues were originally harvested and stored their final use had not been decided. Consequently, more time may have elapsed before they were frozen down, thus allowing RNA degradation to occur, than would have done if a definite application for later RNA extraction had been envisaged.

While human lung tissue was the preferred RNA source for this study, the points raised so far serve to illustrate how difficult lung is to work with. These experiments also highlight the dependence of the scientist on clinicians for the collection of suitable starting materials. Some procedures, such as pneumonectomy or lobectomy, are lengthy and intricate, and are not ideal for obtaining tissue that is ultimately destined for RNA work. Although the surgeons involved were on the whole extremely helpful, their priorities were obviously to the patient and not to the surgical specimen. It was sometimes difficult to convey to non-research, clinical staff the importance of obtaining the unfixed specimen as rapidly as possible. Several pots containing lung tissue were left in the operating theatre for up to several hours before any instructions for their collection were conveyed. On one occasion, a lung adenocarcinoma was even placed in formol-saline fixative by an over-helpful theatre technician! Consequently, several potentially useful samples were inadvertently rendered unusable.
Chapter 4  Adaptation and development of the DDRT-PCR technique for use with human lung RNA

This series of trouble-shooting experiments served to focus the mind on every aspect of the DDRT-PCR method. They also emphatically demonstrate the effects that a range of factors can have on an experiment. Some of these, such as the thickness of a PCR tube, can be controlled, while others, such as the British weather, cannot.
Chapter 5 - Differential mRNA expression between a normal and a lung tumour cell line
Chapter 5 Differential mRNA expression between a normal and a lung tumour cell line

5.1 Introduction

A preliminary DDRT-PCR study using RNA obtained from human lung cell lines was set up. Human lung epithelial cell line L-132 acted as the normal control sample, while the tumour RNA was taken from the pulmonary adenocarcinoma cell line SKLU-1. Various combinations of DDRT-PCR primers were used and the resultant bands were then analysed further.

5.2 Results

5.2.1 DDRT-PCR

RNA samples from normal human lung epithelial cell line L-132 and from human lung adenocarcinoma cell line SKLU-1 were used for DDRT-PCR, and were screened with a total of 20 primer combinations. Primers used were as follows:

<table>
<thead>
<tr>
<th>Arbitrary primers</th>
<th>Anchor primers</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP1</td>
<td>T_{12}MA T_{12}MC T_{12}MG T_{12}MT</td>
</tr>
<tr>
<td>AP2</td>
<td>T_{12}MA T_{12}MC T_{12}MG T_{12}MT</td>
</tr>
<tr>
<td>AP3</td>
<td>T_{12}MA T_{12}MC T_{12}MG T_{12}MT</td>
</tr>
<tr>
<td>AP4</td>
<td>T_{12}MA T_{12}MC T_{12}MG T_{12}MT</td>
</tr>
<tr>
<td>AP5</td>
<td>T_{12}MA T_{12}MC T_{12}MG T_{12}MT</td>
</tr>
</tbody>
</table>

Reproducible, clear differential display patterns were observed in both normal (L-132) and tumour (SKLU-1) cell line RNA samples. Each lane revealed ~100-150 discrete bands. Patterns obtained from duplicate samples were ~95% equivalent and clear differences were present between different primer combinations and between normal and tumour RNA samples. See figure 5.1.
Figure 5.1

An autoradiograph showing the differential banding pattern obtained from RNA samples extracted from human lung cell lines L-132 (normal epithelium, =N) and SKLU-1 (adenocarcinoma, =T).

The differences in banding patterns produced from each template with the AP$_1$/T$_{12}$MN primer combinations can clearly be seen. This PCR was performed three times in total, each time duplicates were set up for all reactions to verify any banding differences observed.

Five of the most intense banding pattern differences were selected for further analysis. These bands are indicated by the numbered bars beside the lanes.

Subsequent sequencing of the cDNA within these bands allowed the identification of the cDNA tags using a BLAST database search. Their identities are shown at the foot of the figure.
Figure 5.1 DDRT-PCR gel autoradiograph from cell line RNA study.

N = L-132 cell line RNA
T = SKLU-1 cell line RNA

T12MA  T12MC  T12MG  T12MT

1/2= Fibronectin cDNA
3= Guanylate binding protein
4/5= human cDNA clone 72720
5.2.2 Cloning and sequencing of the differentially displayed bands

Five candidate bands representing the strongest display differences were analysed further, producing 3 sequence “tags”. Coincidentally, all these bands originated from the same arbitrary primer, AP1. Two of the sequence tags were over-expressed in the tumour RNA and the other was over-expressed in the normal RNA sample, see figure 5.1. Comparison to database sequences suggested their possible identities could be human fibronectin (FN) mRNA (100% similarity- 230/230 bases), human guanylate binding protein B (GBP) mRNA (95% similarity- 63/65 bases) and Homo Sapiens cDNA clone 72720 (c72) (100% similarity-106/106 bases). Partial sequences from each cDNA tag are shown in figure 5.2.

5.2.3 Northern & Southern blotting and probe hybridisation

Initially, northern blots of the cell line RNA samples were prepared and probed with each of the cDNA fragments described above. The probes were prepared by cutting out the cDNA sequence from the cloning vector using restriction enzyme digests using Nsi I. (See Chapter 3, figure 3.1). The freed inserts were then verified by size and purified by agarose gel electrophoresis. No signal was detected despite repeated attempts using various reaction conditions and stringencies. The probes were successful in self-annealing so reaction parameters were altered again in an attempt to optimise conditions. Reactions were set up in the presence and absence of dextran sulphate, at sequentially lower hybridisation temperatures and using smaller amounts of cDNA in the labelling mix to try to maximise labelling efficiency. Unfortunately, results remained disappointing. The final variable to examine was the method in which the probes were made. Successful hybridisation had been achieved using PCR-generated GAPDH probes, so instead of using restriction fragments, the cDNAs probes were prepared by amplification of their plasmid clones using sequencing primers for PCR. These new probes were first tried out on cell line RNA dot blots and this time they worked.

Only the FN probe showed differential hybridisation on both the dot blot and northern membranes, binding to the tumour but not to the normal RNA samples. The GBP probe showed initial mild differential binding on the dot blot, but no differences were seen on the northern membrane and the c72 probe failed to bind to either of the RNAs on the dot blot and was not used in further experiments. See figure 5.3.
Figure 5.2 Examples of Genbank database sequence comparisons of cloned cell line cDNA fragments.

HUMFN2 human fibronectin (FN) 3’ coding region and flank mRNA
(100% match)

Query: CCAAGATGCAAATGTTTGGAAATGATAGACCAAAATTTTAAGT
Sbjct : CCAAGATGCAAATGTTTGGAAATGATAGACCAAAATTTTAAGT
Query: AGGAAAGTCACCCCAACACTTCTGCTTCA
Sbjct : AGGAAAGTCACCCCAACACTTCTGCTTCA

HUMGBP1 human guanylate binding protein isoform 1 (GBP-2) mRNA
(95% match)

Query: CCTGGACATGG[TCAGAGATCCACATGACAGGCCCAATGGT
Sbjct : CCTGGACATGG[TCAGAGATCCACATGACAGGCCCAATGGT
Query: CCTATTGAGAACAACACTAATGGG
Sbjct : CCTATTGAGAACAACACTAATGGG

Homo sapiens cDNA clone 72720 3’ end
(100% match)

Query: CTAAACGTCCTATGAAGAGAGACAGTGTATCTTTTAATTAATT
Sbjct : CTAAACGTCCTATGAAGAGAGACAGTGTATCTTTTAATTAATT
Query: GGCACCACCTGGAAATTTCATTTACCTTACGGCTATTCT
Sbjct : GGCACCACCTGGAAATTTCATTTACCTTACGGCTATTCT

Please note only partial sequence information from each cDNA tag is shown here.
Figure 5.3 (a) and (b)

Autoradiographs of nylon membrane containing a dot blot of RNA from cell lines L-132 (normal epithelium, =N) and SKLU-1 (adenocarcinoma, =S). The membrane was used in hybridisation reactions with probes generated from (a) the fibronectin (FN) cDNA tag and (b) a control probe of GAPDH. [Approximately 2 μg RNA present per dot]

As on the DDRT-PCR gel, the tumour RNA from SKLU-1 demonstrated a higher level of expression of FN than the normal L-132 RNA. No signal was obvious in the latter. Hybridisation of the GAPDH control probe appeared equal in each case.

[The northern blot autoradiograph for these samples can be seen in figure 5.4 (a)]
Figure 5.3 Cell line sample blots hybridising to FN and GADPH probes.

a) RNA dot blot with FN probe

b) RNA dot blot with GAPDH probe

$L = L$-132 cell line RNA (N)
$S = SKLU	ext{-}1$ cell line RNA (Ca)
Figure 5.3 (c)

Autoradiographs of nylon membranes containing RNA from cell lines L-132 (normal epithelium, =N) and SKLU-1 (adenocarcinoma, =Ca).

The left-hand figure represents a dot blot hybridising to the probe generated from the guanylate binding protein (GBP) cDNA tag. A weak signal can be seen where the probe has hybridised to the SKLU-1 RNA. This verified the result obtained from the DDRT-PCR gel where a similar difference in expression was observed. [Approximately 2 μg RNA present per dot]

The northern blot on the right-hand side failed to demonstrate this result conclusively. The numerous marks that are present throughout its surface are merely background signals, although one or two do look deceptively band-like. Lane positions are indicated by L. The “bands” are not actually within the lanes. [Approximately 10 μg RNA present per lane]
Figure 5.3 Hybridisation studies using GBP2 cDNA probe.

C) Dot blot

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N (L-132)</td>
<td>Ca (SKLU-1)</td>
</tr>
</tbody>
</table>

\[\text{Weak signal} \quad \text{Ca > N}\]

C) Northern blot

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N (L-132)</td>
<td>Ca (SKLU-1)</td>
</tr>
</tbody>
</table>

\[\text{Weak signal} \quad \text{Ca > N}\]
Autoradiograph of an RNA dot blot from cell lines L-132 (normal epithelium, =N) and SKLU-1 (adenocarcinoma, =Ca). The membrane was used in a hybridisation reaction with a probe generated from the human c720720 (c72) cDNA tag. [Approximately 2 µg RNA present per dot]

Apart from background spots, no signal was apparent in either set of lanes. Consequently, no northern blot was performed using this probe.
Figure 5.3 Hybridisation studies using c72720 (c72) cDNA probe.

N (L-132)  Ca (SKLU-1)

Dot blot showing absence of hybridisation of c72 probe to cell line RNA samples.
Membranes were also prepared using cDNA from the original bands from which the probes were cloned, in order to verify binding specificity and act as positive controls, and from Southern blots of genomic cell line DNA (digested with BamH1 and EcoR1) to determine whether any differential binding observed was due to gene amplification. Results are summarised in table 5.1 below.

Table 5.1. Summary of hybridisation results from DDRT cell line probes.

<table>
<thead>
<tr>
<th>RNA dot blot</th>
<th>Total RNA</th>
<th>Genomic RNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northern blot</td>
<td>Southern blot</td>
<td></td>
</tr>
<tr>
<td>Fibronectin</td>
<td>Ca +++</td>
<td>Ca +++</td>
</tr>
<tr>
<td></td>
<td>N -</td>
<td>N -</td>
</tr>
<tr>
<td></td>
<td>Band ~ 7.5 kb</td>
<td>Band ~1.1 kb</td>
</tr>
<tr>
<td>GBP</td>
<td>Ca +</td>
<td>Ca -</td>
</tr>
<tr>
<td></td>
<td>N -</td>
<td>N -</td>
</tr>
<tr>
<td>Clone 72720</td>
<td>Ca -</td>
<td>Not done</td>
</tr>
<tr>
<td></td>
<td>N -</td>
<td></td>
</tr>
</tbody>
</table>

Key: Ca= tumour RNA sample (SKLU-1), N= normal RNA sample (L-132)

+++ = strong positive signal  + = mild positive signal  - = no signal

As shown in the table above, the FN probe did show differential expression with a strongly positive signal in the tumour RNA lane and no signal in the normal RNA lane. The cell line RNA samples for these experiments were obtained from L-132 and SKLU-1. To determine whether this result was due to some artefact of cell culture or to a true difference in gene expression due to the neoplastic phenotype, further RNA samples from human lung adenocarcinoma cell lines A-549 and MOR-P lines were also included on a northern blot. Results demonstrated the FN probe bound to both L-132 and SKLU-1 RNA samples as
before, with A-549 giving a similar sized product. Only the tumour cell line MOR-P failed to
give any signal at all with FN. Different amounts of RNA were loaded onto the blots in an
attempt to compensate for any inadvertent irregularities in loading and also to try to reduce the
signal threshold. However, the same pattern of hybridisation was observed for each dilution.

See figure 5.4.

Results from the other 2 cDNA probes, GBP and c72720, imply the bands were false
positives on the DDRT-PCR gel. This could have been caused by contamination during the
RT or PCR step.

5.2.4 Further analysis of FN sequence tag

5.2.4.1 Sequencing of the FN cDNA tag

The full length FN cDNA tag was estimated to be approximately 400 bp long. The
complete tag sequence was obtained from manual and automated sequencing using both T7
forward and M13 reverse primers and is shown in figure 5.5.

AP1 [AGCCAACCA], the complementary T12MN- 3’end [ACA] and the poly(A) consensus
sequence [AAATAA] are all present. This sequence is an exact match to part of human FN
mRNA sequence EMBL/X02761/HSFIB1 (bases 7311-7689). The full length mRNA
transcript is 7.6 kb.
Figure 5.4

Autoradiographs of nylon membranes containing nucleic acid samples from various human lung cell lines hybridising to fibronectin (FN) and GAPDH probes.

(a) Shows a northern blot of RNA from cell lines L-132 (normal, L), SKLU-1 (adenocarcinoma, S), MOR-P (adenocarcinoma, M) and A-549 (adenocarcinoma, A).

Sequentially of smaller amounts of RNA were loaded to try and obtain a clearer signal difference and determine the signal threshold. In each case, the signal from the SKLU-1 lane was much more intense than that from A-549. No signal was observed in any of the L-132 lanes or those lanes containing MOR-P RNA.

The FN signal obtained corresponded to a ~7.5 kb transcript which was the correct length for human FN mRNA.

GAPDH control probe demonstrated approximately equivalent loading for each set of samples, with diminished signal intensity occurring for each dilution.

(b) Shows a Southern blot of genomic DNA extracted from cell lines L-132 (N), SKLU-1 (Ca) and hybridised to a probe generated from the FN cDNA tag. DNA samples were digested with either BamH1 or EcoR1 enzymes. The signal obtained corresponded to a 1.1 kb fragment of DNA. [Approximately 10 μg DNA loaded per lane]

This Southern blot was performed to look for any evidence of gene amplification that may have explained the apparent differences in FN expression that had been observed. No such result was observed and the FN probe hybridised to both sets of genomic DNA.
Figure 5.4. Hybridisation of FN probe to northern & Southern blots

a) Northern blot of cell line samples with both FN and GAPDH probes

- 10 µg RNA
- 5 µg RNA
- 2.5 µg RNA

L = L-132 sample (N)
S = SKLU-1 sample (Ca)
M = MOR-P sample (Ca)
A = A-549 sample (Ca)
LS = length standard

- FN (7.5 Kb)
- GAPDH (1.2 Kb)

b) Southern blot of cell line genomic DNA with FN probe

(N) L-132 DNA
(Ca) SKLU-1 DNA

LS  BamHI  EcoRI  BamHI  EcoRI
Chapter 5  Differential mRNA expression between a normal and a lung tumour cell line

Figure 5.5. Human Fibronectin cDNA Sequence (3’end)

5’TTTGGTTTGG GATCAATAGG AAACCATATG CAGCCAACCA
AGATGCAAAAT GTTTTGAAAT GATATGACCA AAATTTTAAG
TAGGAAAGTC ACCCAAAACAC TTCTGCTTTC ACTTAAGTGT
CTGGCCCCCCA ATACTGTAGG ACAACAGCATG ATCTTGGTAC
TGTGATATTT TAAATATCCA CAGTACTCAC TTTTTCAAAA
TGATCCTAGT AATGGCCTAG AAATATCTTT CTTCTTCTTG
TTATTTATCA ATTTTTTTCA GTATTITTTAT ACGGAAAAAA
TTGTATGGA AACCCTTATG ATGCAGTTGA TAAGAGGAAT
TTGGTGATAT ATGCTGTTGT GATATTTTT TATACGTAT
GTGCCAAAGC TTTACTACTG TGGAAGACA ACTGTTTAAT
AAAAGATTAA CATTCCACA3’

7310
7350
7390
7430
7470
7510
7550
7580
7600
7660
7680

Key:  AGCCAAACCA = AP1
ACA = complementary T12MN-3’end
AAATAA poly(A) consensus sequence
Sequence in italics = cloning vector region
7310 -7680 = base position (according to EMBL sequence readout)

This figure represents the full sequence of the fibronectin cDNA tag isolated from DDRT-PCR using cell line RNA templates. This tag shared 100% homology to the human fibronectin mRNA from the EMBL database. The AP1 strand shown here corresponds to the coding/sense strand sequence.

5.2.4.2 Expression studies using the FN probe

Expression studies were performed with the FN probe using a total RNA northern blot from a range of matched human and mouse tissues. See table 5.2 and figures 5.6 - 5.7. Differential hybridisation of the FN probe was observed in some samples when human tissue RNAs were used. The FN probe bound in varying degrees to most tissues, both normal and tumour. A relative increase in FN probe signal was observed in the tumour RNA samples from brain, kidney and pancreas, also in liver hyperplasia and pseudotumour of the lung. Inequalities in RNA loading, as indicated by the signal from the control probe GAPDH, made
Figure 5.6

Autoradiographs of northern blot membranes containing various matched human tissue RNA samples showing hybridisation of the fibronectin (FN) probe.

(a) Study examining FN expression in a range of matched human tissue RNA samples (N= normal tissue, Ca= tumour tissue).

(b) Corresponding RNA samples hybridising to control probe GAPDH.

Although the FN results appear a little confusing at first glance, if the GAPDH blots are examined and the relative RNA loading is taken into consideration, FN expression appears greater in most of the tumour samples relative to the normal tissues. (This is by visual inspection from the human eye only.)

[(a and b) Left-hand figure represent a home-made blot containing approximately 10 μg RNA per lane, while the right-hand figure represents a blot purchased from Invitrogen and contains approximately 20 μg RNA per lane.]
Figure 5.6 Human tissue RNA samples hybridising with fibronectin (FN) probe.

a) FN probe signal

N Ca N Ca N Ca N Ca N Ca N Ca N Ca

Br. Kid. Liv* Lung*
Ca N Ca N Ca N Ca N Ca N Ca N Ca

Key: Oes = adenocarcinoma (AC) oesophagus, Stom = AC stomach, Colo = AC colon, Ova = AC, ovary, Lung = squamous cell Ca, Panc = AC pancreas, Liv = normal liver, Br = brain astrocytoma, Kid = renal cell carcinoma, Liv* = liver hyperplasia, Lung* = pseudotumour lung, N = RNA from normal tissue, Ca = RNA from tumour tissue

Note: Right-hand=Northern Territory™ membrane purchased from Invitrogen, UK.
Figure 5.7

 Autoradiographs of a northern blot membrane containing RNA from various mouse tissues and hybridised with the fibronectin (FN) probe. RNA was extracted from several organs from a male C57Bl/AJ mouse that had not been treated with any chemical carcinogens.

[Approximately 5 μg RNA loaded per lane]

Hybridisation of the FN probe shows mostly background smearing, probably indicative of genomic DNA contamination, with a weak positive signal in the lung and heart RNA lanes.
Figure 5.7 Mouse RNA samples hybridising with fibronectin (FN) cDNA probe.

Lu = Lung RNA
Li = Liver RNA
Br = Brain RNA
Ht = Heart RNA
interpretation of the remaining samples less straightforward. In the mouse tissue RNA samples, only very weak signal was obtained from the FN probe in any of the samples despite strong signal from the control probe in each lane. See figure 5.7 Why similar results were not obtained with the mouse tissue is unclear. According to the database scans, the FN sequence showed 80% homology with segments of mouse FN, 96% homology with segments of rat FN and 88% homology with segments of rabbit FN.

Table 5.2. Summary of northern blotting data using human tissue.

<table>
<thead>
<tr>
<th>RNA source</th>
<th>GAPDH probe signal</th>
<th>FN probe signal</th>
<th>Relative intensity of FN signal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Ca</td>
<td>N</td>
</tr>
<tr>
<td>Brain astrocytoma</td>
<td>+++</td>
<td>+++</td>
<td>+</td>
</tr>
<tr>
<td>Kidney RCC</td>
<td>+++</td>
<td>+++</td>
<td>+</td>
</tr>
<tr>
<td>Liver hyperplasia</td>
<td>+++</td>
<td>+++</td>
<td>+</td>
</tr>
<tr>
<td>Lung pseudo-Ca.</td>
<td>+++</td>
<td>+++</td>
<td>+</td>
</tr>
<tr>
<td>Oesophagus AC</td>
<td>+++</td>
<td>+++</td>
<td>+</td>
</tr>
<tr>
<td>Stomach AC</td>
<td>+</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>Colon AC</td>
<td>++</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td>Ovary AC (Ca)</td>
<td>(x)</td>
<td>+++</td>
<td>(x)</td>
</tr>
<tr>
<td>Lung SCC</td>
<td>+</td>
<td>+++</td>
<td>+</td>
</tr>
<tr>
<td>Pancreas AC</td>
<td>+/-</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>Liver (N)</td>
<td>+</td>
<td>(x)</td>
<td>+</td>
</tr>
</tbody>
</table>

Key: N = normal RNA sample  
Ca = tumour RNA sample  
AC = adenocarcinoma  
RCC = renal cell carcinoma  
x = not done (sample not present)  
+/-/0 = signal intensity strong/weak/absent  
Ca = not applicable  
pseudo-Ca = pseudotumour  
n/a = not applicable  
? = interpretation equivocal

[Signal intensity was assessed by visual inspection of the autoradiograph only.]
5.3 Discussion

5.3.1 Explanation of experimental results

5.3.1.1 Behaviour of MOR-P cell line with FN probe

The lack of hybridisation of the MOR-P RNA when exposed to the FN probe was not really so unexpected. One explanation could be that the MOR-P originated from a different cell type than the L-132 and A549 lines, i.e. from type 2 alveolar cells instead of Clara cells, or that it was from an adenocarcinoma which had metastasised to the lung rather than from primary lung neoplasm. More details were required on the donor tissue and the general background of this cell line. Attempts were made to identify the origin of MOR-P but little data was available as the original depositor of the line was deceased. Despite extensive enquiries, the only information it was possible to obtain was that MOR-P had been grown from a human pulmonary adenocarcinoma.

5.3.1.2 GBP & c72 results

The failure of the GBP and c72 sequence tags to be verified by hybridisation analysis highlights the main criticism of DDRT-PCR, namely the high rate of false-positive bands obtained. As mentioned in chapter 2, the proportion of false positives is now believed to be as high as 70% [Debouck, 1995, Sunday, 1995]. It can be seen from figure 5.1 that the bands representing GBP and c72 appear quite distinctly in either the normal or tumour RNA lanes, leading the unsuspecting observer to believe they were a fairly “safe bet” for being true expression differences. These results may be due to several factors. It is possible that the bands represent contamination by genomic DNA that may not have been removed during RNA preparation. An alternative explanation is that they are due to co-segregating bands of the same molecular weight, one band overlapping another and resulting in an exaggerated signal intensity. This would also explain why the probes made from the bands did not hybridise to the northern blots. When each tag cDNA was cloned, if the resulting colonies contained several populations with different inserts, either from overlapping bands or from genomic
contamination, any probes made from individual plasmid preps could either hybridise to both normal and tumour RNA samples or hybridise to neither.

### 5.3.1.3 FN expression studies

The observation that the FN probe bound to all intact human tissue RNA samples may be explained by a relative insensitivity of the FN fragment used in the probe, i.e. it may have been that different FN splice variants were being expressed in the normal and tumour samples, which were not able to be discriminated by the cDNA probe used. To conclusively resolve this problem the samples would have required probing or PCR with the specific splice-variant sequences, providing such probes or primers could produce different sized fragments that could be demonstrated on a northern blot or on an agarose gel. Alternatively, tissue sections would have required testing with a panel of antibodies for these same variants. All of these procedures would have required the investment of far more time than was available. It was decided, therefore, not to continue with analysis of this cDNA tag any further and not to attempt *in situ* hybridisation using this sequence.

There seemed to be an inconsistency in the size of band obtained from northern blotting (7.5 kb) and that from Southern blotting (1.1 kb). How could a genomic DNA be smaller than its mRNA transcript? It must be remembered that the genomic DNA used in the Southern blot had been fragmented by enzymatic digestion, and so the 1.1 kb DNA fragment must have contained the target site for the FN cDNA probe.

### 5.3.2 Significance of differential expression of FN in human lung cell lines

What, then, is the significance of changes in the expression of FN between adenocarcinoma and normal lung cell samples and could it have a role in pulmonary adenocarcinoma pathogenesis?

FN is a dimeric extracellular matrix protein that allows cell adhesion and migration. It has become clear that FN is a multi-functional binding protein, possessing the ability to bind fibrin, heparin, collagen and gelatin [Infusa *et al*, 1996]. Cells also bind to FN via membrane receptors of the integrin family, composed of alpha and beta chains. FN is physiologically...
important during embryogenesis, organogenesis and wound healing, and in the pathogenesis of metastasis. Expression of FN increases during fetal development, particularly during type 2 cell differentiation, decreasing thereafter to virtually undetectable levels in adult lung [Dean et al, 1991]. FN consists of 2 subunits each made from 3 types of homologous repeat. The exon of the type III homologies shows alternative splicing of its pre-mRNA, termed ED-A and ED-B [Oyama et al, 1990]. See figure 5.8. Alternative splicing is believed to be regulated in a cell type-specific, tissue-specific and developmental stage-specific manner [Oyama et al, 1990, 1993]. Although 20 splice variants of human FN have been characterised, it remains unclear what function and significance these variants possess [Mardon et al, 1993]. These variants are further divided into two biological classes, namely cellular and plasma FN. Cellular FN contains ED-A(+) and ED-B(+) and is synthesised by many types of cell, whereas plasma FN is produced only in the liver and lacks both ED-A and ED-B regions. Up-regulation of cellular FN has been detected in oxidant-damaged lung tissue, but the same FN subtype has also been demonstrated in normal lung. This raises the question whether damage changes the splicing patterns or merely increases pre-existing levels of the splice variants [Ffrench-Constant, 1995].

Opinions vary on the production of FN by tumour cells. Some studies maintain cancer cells do not produce FN, while others claim that they do [Infusa et al, 1995]. Loss of FN expression from the cell surface has been found to correlate with oncogenic transformation in various cell lines, but the evidence for involvement in tumorigenicity is less convincing [Mardon et al, 1993]. Alteration in splicing towards an embryonic pattern is known to occur in tumour growth, and the hepatocellular carcinoma system has been extensively studied [Tavian et al, 1994]. Increased expression of certain FN splice variants have been documented in several human tumours, including breast, liver, colon and lung [Oyama et al, 1990, 1993, Pujuguet et al, 1996]. The ED-B segment has been shown to be present in the tumour vessels and connective tissues surrounding tumour nodules in several organs [Inoue et al, 1996, Tavian et al, 1994, Oyama et al, 1993]. In the liver, ED-A(+) mRNA detection closely correlates with the degree of invasion [Oyama et al, 1989]. This information implies that ED-A/ B(+) FNs are connected with tumour-associated stroma rather than with the tumour cells themselves, implying a role in tumour progression and invasion [Inoue et al, 1996]. In the lung, increased expression of ED-B mRNA has been demonstrated in all tumour types, including AC [Oyama et al, 1990]. However, ED-A mRNA showed similar levels of expression in normal fetal and adult lung tissue and in lung tumours [Oyama et al, 1990]. It
Figure 5.8. Structure of the fibronectin molecule.

A) Organisation of fibronectin

B) Alternative splicing pattern of EDIIIA pre-mRNA (simplified)

Inclusion of EDIIIA or B exons gives rise to EDIIIA/B+FN. Exclusion gives rise to EDIIIA/B-FN.

[EDIIIB pre-mRNA undergoes the same pattern of splicing]

Re-drawn from: Mardon H.J., et al. 1993 (see References)
appears that tumour cells cause increased production of ED-A(+) FN from surrounding fibroblasts [Inoue, 1996]. The function of ED-A(+) FN remains unknown at present but it may aid the deposition of FN in the tumour stroma. This, in combination with other adhesion-promoting/inhibiting agents, may help in the migration of tumour cells [Inoue, 1996].

Recently, attention has shifted to examine the role of integrin receptors that bind FN. Integrins are heterodimeric complexes composed of an alpha and beta subunit. The $\beta_1$ subfamily combines with one of nine members of the $\alpha$ chain family ($\alpha_{1-9}$). Integrins that bind FN include $\alpha_5\beta_1$, $\alpha_4\beta_1$ and $\alpha_3\beta_1$ [Agrez et al, 1996]. There is increasing evidence that expression of FN receptors is closely associated with malignant transformation. Studies on the adenoma-carcinoma progression in colonic cancer showed progressive loss of $\alpha_5\beta_1$ and $\alpha_3\beta_1$ [Agrez et al, 1996]. In the lung, loss of expression of $\alpha_3\beta_1$ has been reported in both NSCLC and SCLC samples [Bartolazzi et al, 1995, Smythe et al, 1995]. Results obtained for cell lines and tissue samples have been conflicting, and illustrate the complexity of the relationship.

FN is extremely sensitive to proteolysis and fragments of FN have been found to promote cellular adhesion and motility. Solid tumours are known to produce proteolytic enzymes during invasion through the basement membrane. If such cells should encounter partially degraded pools of FN in the extracellular matrix, the fragments could serve to encourage further tumour migration [Margolis et al, 1996]. This group demonstrated that specific fragments of FN digested by proteases were responsible for this stimulation and that these fragments were also able to activate protein kinase C (PKC). He concluded that matrix mediated motility of invasive tumours involves the PKC signal transduction pathway and that a reversal in the malignant phenotype of a tumour could occur via inhibition of this pathway. Multiple tumour types have been prevented from forming metastases using superfibronectin (sFN), a highly adhesive molecule that is a powerful inhibitor of cell migration in vitro [Pasqualini et al, 1996]. Recombinant FN fragments have also been used to inhibit lung metastases [Saiki et al, 1990].

Fibronectin does, therefore, have a role in tumour development. At the earlier stages with alterations in splice-variant expression and at more advanced stages with tumour dissemination. Thus, the altered expression observed from this cell line study may be indicative of changes occurring in vivo during the development of adenocarcinoma. These results
provided great encouragement concerning the use of DDRT-PCR with a lung RNA system. This series of experiments also helped in preparation for the next stage of the study: the use of human lung tissue samples.
Chapter 6 - Differential mRNA expression between matched human lung tumour samples
Chapter 6  Differential mRNA expression between matched human lung tumour samples

6.1 Introduction

Five matched cases of pulmonary adenocarcinoma were used to prepare total RNA. In these cases, normal tissue taken from the lung resection away from the area of the adenocarcinoma could act as a control against the tumour. Tissue specimens were collected from Birmingham Heartlands NHS Trust Hospital, East Birmingham and Glenfield General Hospital, Leicester. For details see Table 6.1, below.

Table 6.1. Matched pulmonary adenocarcinoma cases collected

<table>
<thead>
<tr>
<th>Case details</th>
<th>Specimen collection date</th>
<th>Histological diagnosis</th>
<th>Fate of extracted RNA samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1 (Male/ 50 yrs)</td>
<td>02/95 (BHH)</td>
<td>Adenopapillary carcinoma</td>
<td>RT (K-ras) + DDRT + (API)</td>
</tr>
<tr>
<td>#2 (Male/ 67 yrs)</td>
<td>07/95 (BHH)</td>
<td>Adenocarcinoma</td>
<td>RT (K-ras) + DDRT -</td>
</tr>
<tr>
<td>#3 (Female/ 67 yrs)</td>
<td>02/95 (GGH)</td>
<td>Adenocarcinoma</td>
<td>RT (K-ras) - DDRT -</td>
</tr>
<tr>
<td>#4 (Female/ 55 yrs)</td>
<td>01/95 (GGH)</td>
<td>Adenocarcinoma</td>
<td>RT (K-ras) + DDRT -</td>
</tr>
<tr>
<td>#5 (Male/ 56 yrs)</td>
<td>02/95 (GGH)</td>
<td>Adenocarcinoma</td>
<td>RT (K-ras) + DDRT -</td>
</tr>
</tbody>
</table>

Key: BHH = Birmingham Heartlands NHS Trust Hospital, East Birmingham, GGH = Glenfield General Hospital, Leicester,
RT (K-ras) = control reverse transcription PCR with K-ras primers
DDRT = differential display PCR, API = DDRT arbitrary primer 1
+/- = successful/unsuccessful application

The specimens were collected immediately following surgery and taken as quickly as possible to Histopathology for sectioning and freezing in liquid nitrogen. Upon receipt of the tissue, this entire process could not be done in less than 20 minutes. Approximately 1 cm³ of
tissue was sampled from the tumour, the surrounding normal tissue and distal normal lung. Larger amounts could not be removed as the specimens were primarily intended for clinical diagnostic work and not for research purposes. Great care was taken not to distort the original architecture of the specimen so it would not cause confusion to the duty histopathologist when the organ came to be formally inspected. The remainder of the specimen was immersed in 10% formol saline overnight and embedded in paraffin wax for diagnostic use.

Total RNA was prepared from the tissue samples via the RNAzol™ B method initially. The RNA was then quantified, its quality checked by agarose gel electrophoresis and amplified via RT-PCR with K-\textit{ras} primers. As discussed in Chapter 4, there were some difficulties in obtaining a good differential display signal from many of the RNA samples. Despite the problems encountered with the RNA extracted from the tissues, the samples from test case 1 were good enough to produce a reasonable DDRT-PCR result from which several differentially displayed bands were selected for further analysis.

The patient from which this tissue was taken was a 50 year old Caucasian male. He presented with a history of cough and haemoptysis and was found to have a radiological opacity in the lower lobe of his left lung. There was no past medical history of note other than a myocardial infarction in 1993. The patient had smoked ~20 cigarettes per day during his adult life. A lobectomy was carried out in February, 1995, following which the patient made a good recovery. The lobectomy specimen was sent for histological diagnosis. Examination of fixed tissue sections stained with haemotoxylin and eosin revealed a well-differentiated pulmonary adenocarcinoma.
6.2 Results

6.2.1 DDRT-PCR

The differential display of the RNA samples from case #1 identified various banding pattern differences. Of these, eight were chosen for the study and named H1-8, where H = human tissue cDNA (see figure 6.1). These particular bands were selected as they demonstrated the clearest signal differences between normal and tumour lanes. They were also selected on the basis of length, as cDNA probes above 150 base pairs had a better chance of hybridising on the northern blots than those below this size [BioGene protocol, 1995]. All the bands were obtained from the AP1/ T12MN combination (N=A,C or G). Bands H1, H2 and H6 showed greater expression levels in the tumour RNA sample, while bands H3-5, H7 and H8 were expressed at higher levels in the normal RNA sample. See table 6.2 for details. These combinations represented only 5% of the possible primer couplings. However, as only limited amounts of cDNA were produced and much was used up in the trouble-shooting experiments (described in Chapter 4), it was not possible to perform other primer combinations. In reality, it is neither practical, nor possible, for one person to attempt to analyse all primer combinations using the manual methods described in this project, and the subsequent analysis of these 8 bands provided more than enough work.

Table 6.2. Details of bands obtained from human tissue case #1

<table>
<thead>
<tr>
<th>cDNA band</th>
<th>Primers</th>
<th>Approx. size (bp)</th>
<th>Present in tumour RNA</th>
<th>Present in normal RNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AP1 x T12MA</td>
<td>650</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>2</td>
<td>AP1 x T12MA</td>
<td>350</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>3</td>
<td>AP1 x T12MA</td>
<td>250</td>
<td>NO</td>
<td>YES</td>
</tr>
<tr>
<td>4</td>
<td>AP1 x T12MC</td>
<td>500</td>
<td>NO</td>
<td>YES</td>
</tr>
<tr>
<td>5</td>
<td>AP1 x T12MC</td>
<td>600</td>
<td>NO</td>
<td>YES</td>
</tr>
<tr>
<td>6</td>
<td>AP1 x T12MG</td>
<td>400</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>7</td>
<td>AP1 x T12MG</td>
<td>800</td>
<td>NO</td>
<td>YES</td>
</tr>
<tr>
<td>8</td>
<td>AP1 x T12MG</td>
<td>700</td>
<td>NO</td>
<td>YES</td>
</tr>
</tbody>
</table>
Figure 6.1

An autoradiograph showing the differential banding pattern obtained from RNA samples extracted from matched human adenocarcinoma case #1.

The differences in banding patterns produced from normal (N) and tumour (T) tissue RNA templates with the AP1/ T12MN primer combinations can clearly be seen. This PCR was performed twice and duplicates were set up for all reactions to verify any banding differences observed.

Eight of the most intense banding pattern differences were selected for further analysis. These bands are indicated by the numbered arrows beside the lanes. They were named H (human tissue)1-8.

Subsequent sequencing of the cDNA within these bands allowed the identification of the cDNA tags using a BLAST database search. Their identities are shown at the foot of the figure.
Figure 6.1 DDRT-PCR gel autoradiograph from human tissue RNA study.

N= normal tissue RNA sample
T= tumour RNA sample

1= ERCC5 cDNA clone (excision repair protein)
2= human cDNA clone
3= human cDNA clone
4= human cDNA clone
5= human cDNA clone
6= human cDNA clone
7= influenza cDNA
8= human cDNA clone
6.2.2 Cloning and sequencing of the differentially displayed bands

The eight candidate bands were re-amplified and cloned into the pCR™II TA-Cloning® vector before being partially sequenced, using double-stranded DNA as the template with the T7 promoter and M13-reverse primers. See figure 6.2. This information was then compared to known sequences in the computer databases by performing BLAST and FASTA searches. The degree of sequence similarity for each cDNA tag was variable, ranging from 61-99% over regions of 22-223 bases. See table 6.3 for details, please also note that only the top 3 similarities are shown. The identity of each band was decided by matches that had the highest degree of similarity over the longest lengths of the test sequence.

Table 6.3. Database sequence similarities of H-band partial cDNA tags

<table>
<thead>
<tr>
<th>Band: Accession code:</th>
<th>Sequence identity:</th>
<th>Degree of similarity:</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1 dbjD1630SHUMERCC5</td>
<td>Human ERCC5 mRNA</td>
<td>99% (176/177 bases)</td>
</tr>
<tr>
<td>gbL20046HUMERCC5</td>
<td>Human ERCC5 mRNA</td>
<td>99% (176/177 bases)</td>
</tr>
<tr>
<td>dbjD16306MUSERCC5</td>
<td>Mouse ERCC5 mRNA</td>
<td>74% (90/121 bases)</td>
</tr>
<tr>
<td>H2 embZ84475HS162C6</td>
<td>Human cDNA sequence</td>
<td>81% (26/32 bases)</td>
</tr>
<tr>
<td>gbU88315CELHC37H5</td>
<td>C. elegans cosmid</td>
<td>95% (21/22 bases)</td>
</tr>
<tr>
<td>emblZ81103CEM04G12</td>
<td>C. elegans cosmid</td>
<td>68% (32/47 bases)</td>
</tr>
<tr>
<td>H3 gbT05834T05834</td>
<td>Homo sapiens cDNA</td>
<td>84% (178/211 bases)</td>
</tr>
<tr>
<td>gbH82679H82679</td>
<td>Homo sapiens cDNA</td>
<td>88% (145/163 bases)</td>
</tr>
<tr>
<td>gbH58354H58354</td>
<td>Homo sapiens cDNA</td>
<td>83% (177/211 bases)</td>
</tr>
<tr>
<td>H4 gbL15320HUMNUCPHOF</td>
<td>Homo sapiens cDNA</td>
<td>93% (45/48 bases)</td>
</tr>
<tr>
<td>embLX16934HS6H23</td>
<td>Homo sapiens B23</td>
<td>91% (44/48 bases)</td>
</tr>
<tr>
<td>gbL28699HUMB23A</td>
<td>Homo sapiens B23</td>
<td>91% (44/48 bases)</td>
</tr>
<tr>
<td>H5 gbAC000066HSAC000666</td>
<td>Human BAC clone</td>
<td>78% (60/76 bases)</td>
</tr>
<tr>
<td>emblZ22622HSRPTDNAW</td>
<td>H. sapiens repeat region DNA</td>
<td>75% (60/80 bases)</td>
</tr>
<tr>
<td>emblZ84483HS46H23</td>
<td>Human DNA sequence from PAC</td>
<td>72% (62/86 bases)</td>
</tr>
<tr>
<td>H6 gbT56714T56714</td>
<td>Homo sapiens cDNA clone</td>
<td>86% (132/153 bases)</td>
</tr>
<tr>
<td>gbT82304T82304</td>
<td>Homo sapiens cDNA clone</td>
<td>87% (195/223 bases)</td>
</tr>
<tr>
<td>gbT90251T90251</td>
<td>Homo sapiens cDNA clone</td>
<td>88% (137/155 bases)</td>
</tr>
<tr>
<td>H7 emblX574941FLAHAAD</td>
<td>Influenza A</td>
<td>83% (31/37 bases)</td>
</tr>
<tr>
<td>gbK0992FLAHAMB</td>
<td>Influenza A</td>
<td>83% (31/37 bases)</td>
</tr>
<tr>
<td>gbL09063IAU72666</td>
<td>Swine influenza type A</td>
<td>81% (30/37 bases)</td>
</tr>
<tr>
<td>H8 gbR38560R38560</td>
<td>Homo sapiens cDNA clone</td>
<td>61% (50/81 bases)</td>
</tr>
<tr>
<td>emblF02952HSC1HE082</td>
<td>H. sapiens partial cDNA seq.</td>
<td>76% (33/43 bases)</td>
</tr>
<tr>
<td>dbjC12530C10530</td>
<td>C. elegans cDNA clone</td>
<td>67% (38/56 bases)</td>
</tr>
</tbody>
</table>
Figure 6.2. Partial sequence data on human tissue cDNA clones.

Human sequence H1 (total band size ~ 650 bp)

Forward (198)
GCCATGGAGA AAGAATTTGA GCTACTTGAT AGGGCAAAAAC GAAAAACCCA GAAGAGAGGC
ATAAACAAATA CCTTGAAGA GTCACTCAAGC CTGAAAAAGA AGGGCTTTTC AGATTCTAAA
CGAAAGATAA CATCGGGTGG ATTTTGGGA GACCTGCTCT TCAGAATCAT CTGATGG

Reverse (197)
ATTACAAATG GCTGTCATAA CTAATTATAG AGGATAGATA TTTTTAATT AGGTATTTCCT
TTTTCTCCC CTCGGCAATTC TTGTTTTTCT TCTTTTCTTC CCAACACAG ATCTGGGCGGT
CAGGAGGACCATCTCTCTTT TCCCCTCCATC GTCACTCATA TCACACTAGCT TGCTGGT

Human sequence H2 (total band size ~350 bp)

Forward (53)
TGCAATGGAT ATATGCAAGT GAATAATCC AAATATCCAA ACCAAGATCA GGA

Reverse (153)
GGAAAGAATA CCATCAACAG ATGTGTAAC CTAATTATAG AGGATAGATA TTTTTAATT AGGTATTTCCT
TTTTCTCCC CTCGGCAATTC TTGTTTTTCT TCTTTTCTTC CCAACACAG ATCTGGGCGGT
CAGGAGGACCATCTCTCTTT TCCCCTCCATC GTCACTCATA TCACACTAGCT TGCTGGT

Human sequence H3 (total band size ~250 bp)

Forward (121)
ACCAGCCTGA CAAACATGGT GAAACCCCAT CTCCTACTCC ATGCACACAT GTAGGCGGT
TAGAGGTGG TGCACTAGC CTGCAGATAC ATGCATAGTGC CACGAGCCAT T

Reverse (205)
AGGCTGGAGT GCAAACAGCG ATCTACAGCT CATACAATCC TTTGGACCTAC CCAGGATTAC
TACAGCTCAG TCCCAGCCAC AAATGCACAA AAAGTAAATT CAGAATGGAA AAGACTCAAA ACCATCATCA
ACACCAAGAT CAAAAGGACA AGAATCTCCT CAAAAAACAG GAAAAACCTC TCTAAAACAC
C AAAAGGACT ATGGTC

Human sequence H4 (total band size ~500 bp)

Forward (61)
TGCGAATTAGA ACCCGGACAA CATTATCAA ACATGGTAGG GAAAGTCCA CTCTGCACTTT A

Reverse (136)
CTCCAGCCAA AAATGCACAA AAAGTCAATT CAGAATGGAA AAGACTCAAA ACCATCATCA
ACACCAAGAT CAAAAGGACA AGAATCTCCT CAAAAAACAG GAAAAACCTC TCTAAAACAC
C AAAAGGACT ATGGTC
Human sequence H5 (total band size ~600 bp)

*Forward* (154)

TGGTTTCTATA ATTTAGAAA GAGATTTGCC TTTAAAGATG TCAATATAAG AGAATAAGCA
GCCCTATCACA ACCAAAGGTC AGCAAGATAAA TCTCTGGACA CTTTTAAGAA AATCATTGAG
GGCTAGTGCA GTGGCTCACCA CCTATATTCA GCAC

*Reverse* (191)

TCTCGAATTC CTGGCCTCAT CTGCCTAACCT TCTGGCCGTT GGTGCGACGC ACCTGTGACA ATCTGGATCT
GGGAGGAGTTA AATTCATGGA ATTATTTTTA ATTGTAATCT TGCGAATGTA GCAGGAA

Human sequence H6 (total band size ~400 bp)

*Forward* (163)

CAAACACAA AAATTAGCCA GGCTGGTGTT GGTGGCAGGC ACCTGTAAAC CCAGCTACTT
GGGAGGATGA GGCATGAGAA TCGCTTGAAG CCAGGAGGTT GGAGGTTGCA GCAGCCAAGA
TTATACCCCA ACAAACAGCC CTTGGGCAACA GAGCAAGATC TGT

*Reverse* (137)

GGTGCTATAA GCAGGCCTTGC CTGCCACCTC GCCTCAGCTC ATCTGGAGTT TACAGGCA

Human sequence H7 (total band size ~800 bp)

*Forward* (90)

CCTCACTGAT TAAGCATTGG TAACTGTCAG ACCAAGTTTA CTCATATATC TTTAGATTGA
TTGATTTAAA AATCATCTTAA TAAATTA

*Reverse* (134)

CCAGGGGGATG CTAAGGAAAA CTGCTAACCA GGGCTTGATA TGTTTCACTA CTTGCGTTAA
TACAGGCTAC AAAAAAGAAAA GTGGTGGGCA GCCTCAGCTT CACAGGTTT
AGTGAGGCTT GACA

Human sequence H8 (total band size ~700 bp)

*Forward* (117)

AAATTGCGGTT TTAACCTACA CATCAATGTT TTTCCACAGT GCCCCTAAGA AGACTGCTCA
TGGTCAACTA CAAGTGAGAG GAGAAAAAAA AAAGGGAGGTT TTTGGCGCAA AGGACCA

*Reverse* (243)

TTCAAGGCTA AGTGGTTATGA TTGTTATCTCA AGTTCCTAACG ACAGCTATCC ATGTGGTTAT
TCAATACCC TACCAAGATT CATGCTTTCAT TTTTCTGTTT TAAATCAGGC
AGTATTACCT TATATTAGAA TACTATTGGA ACTTACTTGC TTTATCTTAA ACCATGGAGA
TAGCAGGCTG TTGAAATCAA GTGAGATATC AGGAGATTG AGATGAGTCT CTGTATACAG
ATT

Key:  *Forward* = T7 promoter  
       *Reverse* = M13 reverse primer

[Partial 5' and 3' sequences from the cDNA tags are shown which do not overlap.]
6.2.3 Northern & Southern blotting and probe hybridisation

6.2.3.1 Preliminary investigations of differential expression

Tissue preparations yielded insufficient RNA for northern and dot blot analyses, therefore RNA samples from the cell lines L-132 and SKLU-1 were substituted for this purpose. The tissue cDNA probes were then hybridised against the cell line RNA blots. Obviously, this was not an ideal situation but there was little other choice under the circumstances. In order to maximise use of time and labour, all cDNA probes were first screened for any signs of differential expression using an RNA dot blot, and those that gave a positive result were then used with cell line northern blots. Genomic DNA from cell line L-132 was also extracted and digested with restriction enzymes BamH1 and EcoR1 for use in Southern blots. Results are shown in figures 6.3-6.4 and table 6.4. Preliminary screening with the dot blots eliminated the probe sequence from H2, as it bound to both normal and tumour RNA samples. Probe H7 was excluded for similar reasons. The remaining probes were then hybridised with northern blots. Probes H1 and H4 hybridised to both normal and tumour RNA samples and were eliminated from the study. Probes H3, H5 and H6 gave smearing along the lanes of the northern membrane. Only probe H8 showed differential hybridisation on the northern blot, figure 6.5a. When this probe was applied to a genomic Southern blot it bound to the normal DNA (BamH1 digest fragments) but not to the tumour DNA samples, figure 6.5b.

6.2.3.2 Probe competition assay with H3, H5 and H6

Following the elimination of sequences H1, H2, H4 and H7, further tests were carried out on probes H3, H5 and H6 to try to remove or reduce the smearing observed. As the smearing was seen to occur on both RNA and DNA blots, the problem could have been due to several factors. These included degradation of the cDNA used to make the probes, the presence of repetitive elements in the probe sequence, contamination of the probes due to inadequate purification, insufficient blocking DNA in the hybridisation solutions or insufficient stringency of the post-hybridisation washes. Purification methods, blocking DNA and stringency conditions were all satisfactory for probes H1, H2, H4, H7, and H8, thus there seemed no reason to suppose that these variables should suddenly have deteriorated when using probes H3, H5 and H6. All three probes were re-made, as previously described in the
Of the 8 sequence tags:-

H2 & H7 were eliminated for non-specific binding to the dot blot.

H1 & H4 hybridised to both lanes of the northern blot.
H3, H5 & H6 produced non-specific binding caused on both lanes of the northern & Southern blot.

This left only the sequence tag from band H8 to be analysed further.
Figure 6.4 (a)

Autoradiographs of RNA dot blots from cell lines L-132 (normal epithelium, =N) and SKLU-1 (adenocarcinoma, =Ca) demonstrating hybridisation with some of the probes generated from the human tissue cDNA tags. [Approximately 2 μg RNA present per dot]

The results expected (DDRT-PCR gel in figure 6.1) were compared with the results obtained. Probe H2 gave signal intensities that were the reverse of what had been expected, while probe H7 bound with equal intensity to both samples. The remainder produced results consistent with those observed on the original DDRT-PCR gel.
Figure 6.4 Dot blot hybridisation results using cDNA probes H2, 3, 5, 6 & 7.

Expected result (DDRT gel):
- Ca > N
- N > Ca
- N > Ca
- Ca > N
- N > Ca

Observed result (Dot blot):
- N = Ca
- N > Ca
- N > Ca
- Ca > N
- N = Ca
Figure 6.4 (b)

Autoradiographs of nylon membranes containing nucleic acid samples from cell lines L-132 (normal epithelium, =N) and SKLU-1 (adenocarcinoma, =Ca) hybridising with additional probes generated from the human tissue cDNA tags. [Approximately 10 μg RNA loaded per lane.]

Probe H1 demonstrated equal hybridisation signals in both the N and Ca RNA lanes for both northern and Southern blots. Probe H4 demonstrated approximately equal expression levels in the northern blot but did reveal a difference in hybridisation signal levels in the Southern blot. The signal was very faint, however.
Figure 6.4 Hybridisation results with H1 and H4 cDNA probes.

b) H1 probe

Northern blot

Southern blot

N = L-132 DNA/RNA

Ca = SKLU-1 DNA/RNA

H4 probe

Northern blot

Southern blot

N = L-132 DNA/RNA

Ca = SKLU-1 DNA/RNA
(i) Demonstrates the smearing obtained from probe H3. This smearing may have represented genomic DNA or ribosomal RNA contamination of the RNA samples loaded onto the blot. However, this membrane was used with other human probes and did not show the same results. Alternatively, the contamination could have been within the probe itself or due to its binding to a repeat sequence. Consequently, the probe was re-made and a competition assay performed.

(ii) Here, the results of the competition assay can be seen. Unlabelled, sheared, genomic DNA from cell line L-132 was incubated with the probe solution in an attempt to cause annealing of any common repeat elements. This should have left only non-repeat sequence available for hybridisation. Although the signal was marginally cleaner, this strategy made no significant difference to the results.
Figure 6.4. Example of diffuse smearing of cDNA-probe hybridisation signal.

c)

i) Northern blot, x H3 probe

Here, the probe H3 has produced marked smearing on a northern blot. Similar results were observed for probes H5 & H6.

ii) Northern blot, x H3 probe following competition assay

The background smearing has persisted, despite attempts to block repeat elements with unlabelled, sheared, genomic DNA.
### Table 6.4. Summary of hybridisation results from DDRT human tissue probes.

<table>
<thead>
<tr>
<th>Probe</th>
<th>Sequence match</th>
<th>Total RNA dot blot</th>
<th>Total RNA northern blot</th>
<th>Southern blot - BamH1/EcoR1 digests</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1</td>
<td>ERCC5 cDNA</td>
<td>N ++</td>
<td>N = Ca</td>
<td>N = Ca</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ca +</td>
<td>Band ~4 kb</td>
<td></td>
</tr>
<tr>
<td>H2</td>
<td>cDNA clone</td>
<td>N ++</td>
<td>Not done</td>
<td>Not done</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ca ++</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H3</td>
<td>cDNA clone</td>
<td>N +++</td>
<td>N smear</td>
<td>N smear</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ca +</td>
<td>Ca smear</td>
<td>Ca smear</td>
</tr>
<tr>
<td>H4</td>
<td>cDNA clone</td>
<td>N ++</td>
<td>N = Ca</td>
<td>N = Ca</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ca +</td>
<td>Band ~1 kb</td>
<td></td>
</tr>
<tr>
<td>H5</td>
<td>cDNA clone</td>
<td>N +++</td>
<td>N smear</td>
<td>N smear</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ca +</td>
<td>Ca smear</td>
<td>Ca smear</td>
</tr>
<tr>
<td>H6</td>
<td>cDNA clone</td>
<td>N +++</td>
<td>N smear</td>
<td>N smear</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ca +</td>
<td>Ca smear</td>
<td>Ca smear</td>
</tr>
<tr>
<td>H7</td>
<td>Influenza cDNA</td>
<td>N ++</td>
<td>Not done</td>
<td>Not done</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ca ++</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H8</td>
<td>cDNA clone</td>
<td>N +++</td>
<td>N ++</td>
<td>N ++</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ca +</td>
<td>Ca +</td>
<td>Ca -</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Band ~1.9 kb</td>
<td>Band ~2.1 kb</td>
</tr>
</tbody>
</table>

**Key:** Ca = tumour RNA sample, N = normal RNA sample  
+++ = strong positive signal  ++ = moderate positive signal  + = weak positive signal
Figure 6.5 (a)

Sections from an autoradiograph of a northern blot containing RNA from cell lines L-132 (normal epithelium, =N) and SKLU-1 (adenocarcinoma, =Ca) demonstrating hybridisation with the probe generated from the human tissue cDNA tag #8 (H8).

Serial dilutions of template RNA were loaded onto the northern to try to reduce the signal and maximise the intensity differences observed. It can be seen purely from visual inspection that GAPDH signal indicates approximately even loading between pairs of RNA at the same dilution. However, the signal from the H8 probe shows a diminished intensity in all of the tumour RNA lanes. This was later verified by computerised densitometry image analysis software (see text), and was taken as evidence that H8 showed differential expression in N and Ca lung RNA samples.
Figure 6.5a Northern blot hybridisation with H8 and GAPDH probes.

L = L-132 cell line RNA
S = SKLU-1 cell line RNA

Densitometry measurements were made which revealed a 2-3 fold increase in signal of H8 in the L-132 samples compared with the SKLU-1 samples.
Figure 6.5 (b)

Autoradiographs of a Southern blot containing DNA extracted from cell lines L-132 (normal epithelium, =N) and SKLU-1 (adenocarcinoma, =Ca), and demonstrating hybridisation with the probe generated from the human tissue cDNA tag #8 (H8). [Approximately 10 μg DNA loaded per lane.]

The H8 probe did not appear to hybridise to either of the Ca DNA lanes. This could be indicative of a loss of heterozygosity in the tumour for the H8 gene. The band produced corresponds to ~ 2.1 kb.
Figure 6.5b  Southern blot hybridisation with H8 cDNA probe.

(N) L132 DNA  (Ca) SKLU-1 DNA

\[ \text{BamH1} \quad \text{EcoRI} \quad \text{BamH1} \quad \text{EcoRI} \]

H8

GAPDH
Methods section, from freshly produced cDNAs and hybridisation was repeated using these new probes within 24-48 hours of cDNA production. The results remained unchanged with the smearing persisting in each case. A final attempt was made to produce a cleaner signal by blocking binding to any repetitive elements that may have been present within the cDNA sequence, such as the Alu repeat which was present in H5. This was achieved by performing a competition assay, using an excess of unlabelled ("cold") genomic DNA extracted from the L-132 (non-tumour) cell line to out-compete any internal repeats within the probe sequences. The L-132 DNA was sheared into fragments ~500 bp long by exposure to microwaves and added to labelled, purified probe cDNA in an excess of 20:1. The samples were then boiled together to denature and left to anneal at 65°C for 2 hours before being added to the membrane, under the same hybridisation conditions as previously described in the Methods section. Again, smearing in all lanes was observed for each case, and it was concluded that the probes were non-specifically binding, false-positive sequences. To ensure nothing had been missed, each of these cDNA tags was sequenced further to try to determine if they did contain any repetitive elements that could be responsible for the non-specific signals observed. Further sequence analysis showed that H3 and H6 did indeed contain Alu and L1 repeat elements, respectively. See figure 6.6. Consequently, no further action was taken using these probes.

6.2.3.3 Densitometry analysis of H8 sequence tag

Attention was turned to the final probe sequence remaining, H8. Results from the northern blot using L-132 and SKLU-1 RNA samples demonstrated a greater signal in the lanes containing L-132 RNA. A second northern membrane was prepared using decreasing amounts of RNA (10, 5 and 2.5 μg). This was to try and find a point where the signal from the tumour cell line RNA could be “diluted out” relative to the signal from the normal cell line RNA, as the larger quantities used in the original blot could have saturated the signal. Results have been shown in figure 6.5a. Densitometry readings were taken using this northern blot by comparing signal intensities from the resulting autoradiographs with H8 and GAPDH probes. This was performed using ImageQuant© Molecular Dynamics software. The GAPDH signal acts as a standard and is used to normalise the RNA loading in each lane. The relative intensity of each track containing H8 probe signal is then calculated as a ratio against the GAPDH signal. These ratios can then be compared between RNA samples to assess relative expression.
Figure 6.6. Sequence data for H3 & H6 showing the presence of repeat elements

H3 >gb/T05834/T05834 EST03723 Homo sapiens cDNA clone HFBDH18 similar to EST containing Alu repeat. Length = 363

Score = 762 (210.6 bits), Expect = 2.1e-63, Sum P(2) = 2.1e-63
Identities = 178/211 (84%), Positives = 178/211 (84%), Strand = Minus / Plus

Query: 211

TTTTTTTTTTTTTTGACACAGTCTCGTCTTTTTCGCCCAGGCTGGAGNGCAATAGGGCATGAT 152
Sbjct: 4

TTTTTTTTTTTGGAGACAGATCTGCTGTCGCCGCCAGGCTGGAGTGGATAGGTGTGGCAT 63

Query: 151

CTCAGCTCATATACACCTTCGCTCCAGGTTCAAGCAATTCCTGTCGCTAGGCTCCGG 92
Sbjct: 64

CTCAGCTCATATACACCTTCGCTCCAGGTTCAAGCAATTCCTGTCGCTAGGCTCCGG 123

Query: 91

AGTAGCTGGATTATAGGCGTGCACCACCACGCCCAGCTAGTTTTTGATTTTTTAGTAGA 32
Sbjct: 124

AGTAGCTGGATTATAGGCGTGCACCACCACGCCCAGCTAGTTTTTGATTTTTTAGTAGA 183

Query: 31

GATGGGTTTCAGCATGTTTTTCAGCTGGGT 1
Sbjct: 184

GATGAGGTTTCAGCATGTTTTTCAGCTGGGT 214

H6 >gb/W37681/W37681 zc10c12.r1 Homo sapiens cDNAclone 321910 5' similar to contains L1.t2 L1 repetitive element :.

Length = 398

Score = 625 (172.7 bits), Expect = 1.3e-67, Sum P(2) = 1.3e-67
Identities = 137/155 (88%), Positives = 137/155 (88%), Strand = Plus / Plus

Query: 3

CCAGCGCAAACATCATAATGACAGGATCGACACATACATATTTTACCTCAACCTGACACTATTAAATGTG 62
Sbjct: 166

CCAGCGCAAACATCATAATGACAGGATCGACACATACATATTTTACCTCAACCTGACACTATTAAATGTG 225

Query: 63

AATGGGCTAAATGCCCCAATTAAGAGGCACAGACTGGCAAGTTGGATAAAGATAGAGC 122
Sbjct: 226

AATGGGCTAAATGCCCCAATTAAGAGGCACAGACTGGCAAGTTGGATAAAGATAGAGC 285
Chapter 6  Differential mRNA expression between matched human lung tumour samples

of the H8 target sequence. Densitometry readings between L-132 and SKLU-1 demonstrated levels of H8 were 1.4-3.4 higher in L-132 than those in SKLU-1. See table 6.5.

Table 6.5. Densitometry readings for H8 and GAPDH probes

<table>
<thead>
<tr>
<th>RNA sample</th>
<th>RNA loading (µg)</th>
<th>% volume</th>
<th>% volume</th>
<th>Ratio A:B</th>
<th>Final ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>H8 probe (A)</td>
<td>GAPDH probe (B)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L-132</td>
<td>10</td>
<td>22.0</td>
<td>17.7</td>
<td>1.24 (i)</td>
<td>i:iv 3.44</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>21.1</td>
<td>17.0</td>
<td>1.24 (ii)</td>
<td>ii:v 2.58</td>
</tr>
<tr>
<td></td>
<td>2.5</td>
<td>18.7</td>
<td>15.7</td>
<td>1.19 (iii)</td>
<td>iii:vi 1.38</td>
</tr>
<tr>
<td>SKLU-1</td>
<td>10</td>
<td>7.1</td>
<td>19.5</td>
<td>0.36 (iv)</td>
<td>iv:i 0.29</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>6.7</td>
<td>14.0</td>
<td>0.48 (v)</td>
<td>v:ii 0.38</td>
</tr>
<tr>
<td></td>
<td>2.5</td>
<td>5.4</td>
<td>6.2</td>
<td>0.86 (vi)</td>
<td>vi:iii 0.72</td>
</tr>
</tbody>
</table>

Key:  L-132 = normal cell line RNA
       SKLU-1 = tumour cell line RNA

% volume = measurement of signal intensity within area of probe binding on autoradiograph

Work was then undertaken to further characterise the H8 sequence, to determine its range of expression in other tissues and also to ascertain in which lung cell population it was being expressed.
6.3 Discussion

6.3.1 Explanation of experimental results

6.3.1.1 Problems of 3’-UTR sequence tags

The fact that most of the human DDRT-PCR cDNA tag sequences were cDNA clones of unknown identity was frustrating. However, in many cases the cDNAs generated by DDRT-PCR contain sequences that are entirely unknown to the databases. This phenomenon arises because priming is targeted to the region adjacent to the poly(A) tail. As the sequence tags obtained are commonly in the order of 100-500 bp in length, many of the DDRT-PCR products correspond to the 3’ untranslated region (UTR) of the mRNA. Problems occur since this region is frequently not included in sequence databases. The 3’ UTR also shows great inter-organism variation and, in contrast to coding regions, homologous genes may demonstrate dramatic differences in their 3’ UTR sequences [Sompayrac et al, 1995].

This problem can be overcome in several ways. Firstly, PCR conditions can be adjusted to promote the formation of longer cDNA tags [Averboukh et al, 1996]. This is achieved by increasing the dNTP concentration to 20 μM in combination with standard PCR conditions, or by keeping normal dNTP concentrations and increasing the extension time to 90 seconds. Averboukh found that the latter approach maintained a better overall PCR fidelity and specificity. The cDNA tags produced were more than 1 kb in length, thus necessitating their resolution with a 4.5% polyacrylamide gel rather than the usual 6% gel. Secondly, a technique called “inverse-PCR walking” can be applied to determine more of the unknown sequence at the 5’ region of the cDNA tag [Somapyrac et al, 1995]. This method uses known 5’ end sequence to design a primer from which a single-stranded copy of the mRNA is produced via reverse transcription (RT). The cDNA second strand copy is also made, the ends of this double-stranded cDNA are then ligated to give a circle. PCR primers, also designed from the known 5’ region adjacent to the RT primer, are used to amplify the sequences that lie 5’ to the known sequence. The PCR product is then cloned and sequenced, and the additional sequence information is used to design a new set of primers to walk further into the unknown 5’ region. See figure 6.7. Another PCR-based method that also allows additional sequence information to be determined from cDNA tags is the rapid amplification of cDNA ends (RACE-PCR), also known as anchor PCR [Aapte and Siebert, 1993] This procedure allows
Figure 6.7 Strategies used to obtain additional 5' sequence from a DDRT-PCR cDNA tag.

a) Inverse-PCR "walking"

i] Known sequence within the 5' region is used to design primers for reverse transcription (RT) and PCR (P₁ & P₂).

\[ \begin{array}{c}
\text{5'} \\
P₂ \quad P₁ \quad \text{RT} \\
\end{array} \]

ii] Double-stranded cDNA copy of mRNA is produced

\[
\text{cDNA} \quad \text{copy of } \text{mRNA}
\]

iii] cDNA is circularised by ligation

\[
\text{cDNA} \quad \text{is circularised}
\]

iv] PCR primers P₁ & P₂ are used to amplify unknown 5' region. PCR product is then cloned and sequenced. New primers are designed to walk further towards 5' end...

\[ \begin{array}{c}
P₁ \\
P₂ \\
\end{array} \]
b) **Rapid Amplification of cDNA Ends (RACE) PCR**

i] Following cDNA second strand synthesis, an adaptor primer (AdP) is ligated onto the ends of the double-stranded (ds) cDNA molecule

\[
\begin{array}{c}
5' \\
\text{AdP} \\
\text{SSP} \\
\text{ds-cDNA molecule} \\
3'
\end{array}
\]

ii] 5' or 3' RACE is then performed using a sequence specific primer (SSP) designed from known sequence information and a primer sequence complementary to AdP (AdP₁)

(5' RACE only is shown)

\[
\begin{array}{c}
\text{AdP₁} \\
\text{PCR product} \\
\text{SSP}
\end{array}
\]

iii] Full length cDNAs are generated by end-to-end amplification using both 5' & 3' sequence specific primers OR by cutting the 5' & 3' RACE products with restriction enzymes and ligating the fragments into a cloning vector \(\rightarrow\) sequencing of DNA
amplification of sequences between a known region of the target mRNA and the unknown sequence at either the 5' or 3' end. As in the previous method, RACE-PCR uses an RT primer designed from known sequence, but this time it is used in conjunction with a 5' or 3' anchor primer that has been ligated onto the template molecule [Aapte and Seibert, 1993, Chenchik et al, 1996]. If the template mRNA is small, < 2 kb, it may be possible to obtain the full length cDNA sequence from this method without having to screen a cDNA library. See figure 6.7.

6.3.1.2 Why the use of primer $T_{12}MT$ was discontinued

Subsequent to their original report, the originators of the DDRT-PCT technique, Peng Liang and Arthur Pardee, indicated that they no longer recommended the use of the $T_{12}MT$ anchor primer [Liang et al, 1994 and 1995]. Two reasons were cited. The first was that the use of anchor primers with 2 base-redundancy was unnecessary for efficient differential display. Single base-degeneracy was found to be sufficient, thus only anchor primers with a 3'-A,-C or-G needed to be used. Secondly, with two-base degenerate anchors, the $T_{12}MT$ primer was frequently found to cause non-specific smearing. It was for the latter reason that this primer was excluded during these investigations.

6.3.2 Why use human lung tissue?

Given the problems encountered in this project alone when attempting to use human lung tissue, why bother? Why not just use primary tissue cultures, or samples isolated from the micro-dissection of frozen sections? Is the use of RNA extracted from pieces of human lung tissue suitable for this application? It contains not just the tumour population and normal cell-counterparts, but also vascular tissue, neural structures, reticulo-endothelial cell aggregates and various connective tissues. When studying a disease for which there is no exact animal model, it is essential to include human tissue specimens. Information gained by the use of suitable animal models, or from cell cultures, is undoubtedly important in providing understanding of a disease, but has limitations. How far can the behaviour of a pulmonary adenoma in a mouse, or even of human adenocarcinoma SKLU-1 cells in a culture flask, be extrapolated to the behaviour of a pulmonary adenocarcinoma in a human being? An adenocarcinoma is not just a non-active mass of neoplastic cells. It is a dynamic entity.
continuously interacting with its surrounding environment. This environment contains blood vessels to supply it with nutrients, connective tissue structures to support its mass, neural bundles and collections of lymphoid tissue, any of which may be infiltrated by the tumour. What contributions are made by any of these structures during tumourigenesis, either as a cause or an effect? How can such changes be investigated without looking in the very place in which they will occur? At the very least, if these elements have no role in tumour formation and are merely "contaminants", they will be present in both the tumour and its matched normal RNA and thus "cancel out" each other's effect in any subsequent differential display pattern.

The use of primary tissue cultures for lung is possible but there is the problem of an overgrowth of undesirable cell types, such as fibroblasts. The use of established cell lines has provided a valuable source of experimental material (see Chapter 5), but how can we be sure that transformation and/or serial passaging has not caused changes in genotype or phenotype to diminish the similarity to the original source?

The procurement of suitable human lung tissue during this project was not easy and even when it was obtained, it was not the most user-friendly tissue to work with. Perhaps, with the benefit of hindsight, it would have been more prudent to use combinations of several of these approaches, i.e. cell/primary tissue cultures in conjunction with micro-dissected specimens. If the RNA degradation problem could have been sorted out, the yield of total RNA extracted from micro-dissected frozen tissue samples may have been sufficient for use with DDRT-PCR. RNA extraction from the micro-dissection of paraffin embedded tissue has also been described [Mies, 1995, De Andres et al, 1995]. This would give the added advantage of making hundreds of archival samples available for study, but also has the inherent disadvantage of the inevitable degradation of some mRNA species during fixation. Such a comparison of mRNA populations by DDRT-PCR would therefore only be partial at best, and there would be no way of knowing whether the lost mRNA species were significant.

Despite some of the problems incurred during these experiments, the use of human tissue RNA provided the best opportunity to test the experimental hypothesis. From five initial matched cases, only one yielded any results from DDRT-PCR. Of the eight bands chosen for analysis, only one demonstrated differential expression on northern blots. This degree of "redundancy" is normal for DDRT-PCR based experiments. For example, in the study by Chen et al [1996] using DDRT-PCR with human lung cancer RNA samples, 30 bands
were isolated and only one showed differential expression with northern blot analysis. DDRT-PCR is fairly labour-intensive and repetitive in the initial stages, but the interesting part comes once a differentially expressed cDNA tag has been identified. A series of experiments was then devised to examine the H8 cDNA in more detail.
Chapter 7 - Further analysis of the cDNA band H8
7.1 Introduction

Of the eight differentially expressed cDNA tags obtained from case #1, all but one had been ruled out as a false-positive signal. Band H8 was the only one to indicate differential expression during northern blot analysis. (See Chapter 6.) A series of experiments was initiated to learn more about this cDNA. These included full sequence analysis of the cDNA tag, further expression studies and the elucidation of the full cDNA sequence using RACE-PCR.

7.2 Results

7.2.1 Full sequence analysis of cDNA tag

Partial sequence determination had already been carried out using manual sequencing with sub-cloned plasmid DNA containing the H8 insert. (See Chapter 6.) During this experiment 117 bases from the forward strand (AP1) and 243 bases from the reverse strand (T12MN) had been determined using T7 promoter and M13 reverse primers, respectively. (See figure 6.2, Chapter 6.) Estimates from agarose and acrylamide gel electrophoresis put the length of the H8 cDNA band at around 700 base pairs. Thus, approximately 580 bases from the forward strand and 460 bases from the reverse strand were still to be sequenced.

7.2.1.1 Manual sequencing results

Manual sequencing was again employed to determine more of the missing sequence. T7 promoter and M13 reverse primers were used again, but the sequencing gels were allowed to run for a much longer duration in order to maximise the sequence information obtained—this varied from 3.5 to 6 hours. Only a small amount of useful additional information was obtained from this procedure. Thus, internal sequencing primers were designed for each strand from the regions which flanked the unknown middle section. See figure 7.1a. The intention was to use these primers to “walk” along into this missing segment of the cDNA and to extend the lengths of known sequence until a region of overlap between forward and reverse strands was
Figure 7.1: Preliminary sequencing results for H8 template DNA.

a) Manual sequencing using double-stranded H8/pCRII template

i) Forward strand [AP₁] x T7 promoter

\[ AP₁ \]
\[
5' - \text{AGCCACGCAA} \quad \text{TATGGTTTTT} \quad \text{AACCTACACA} \quad \text{TCAATGTTTTT} \quad \text{TCCACAGTGC} \\
\text{CCCTAAGAAG} \quad \text{ACTGCTCATG} \quad \text{GTTCAACTCA} \quad \text{AGTGAAGAGGA} \quad \text{GAAAAA...} \\
\text{AGGAGGATTT} \quad \text{TGGCGAGGAC} \quad \text{CA} \quad \ldots \ldots \quad =125 \text{bases}
\]

ii) Reverse strand [T₁₂MN] x M13 reverse primer

\[ T₁₂MN \]
\[
5' - (\text{T₁₂MN}) \quad \text{TTCAAGGTGA} \quad \text{AGTGGTTATG} \quad \text{ATTTATTTCC} \quad \text{AGTTTTAAC} \\
\text{CAGCTCAGTA} \quad \text{ATCTAGCTAA} \quad \text{TCCATGTTAC} \quad \text{TCAATGTTTTT} \quad \text{TCCACAGTGC} \\
\text{ACTAGAAGAA} \quad \text{CCCCCCCCC} \quad \text{CCCCCCCCC} \quad \text{CCCCCCCCC} \quad \text{CTGTGTTTAC} \\
\text{TACCGCGGGG} \quad \text{AGTGGTTATG} \quad \text{ATTTATTTCC} \quad \text{TCAATGTTTTT} \quad \text{TCCACAGTGC} \\
\text{GATT} \ldots \ldots \quad = 244 \text{bases}
\]

b) Automated cycle sequencing using H8/pCRII template

i) Forward strand [AP₁] x T7 promoter

\[ AP₁ \]
\[
5' - \text{AGCCACGCAA} \quad \text{TATGGTTTTT} \quad \text{AACCTACACA} \quad \text{TCAATGTTTTT} \quad \text{TCCACAGTGC} \\
\text{CCCTAAGAAG} \quad \text{ACTGCTCATG} \quad \text{GTTCAACTCA} \quad \text{AGTGAAGAGGA} \quad \text{GAAAAA...} \\
\text{AGGAGGATTT} \quad \text{TGGCGAGGAC} \quad \text{CA} \quad \ldots \ldots \quad =125 \text{bases}
\]

ii) Reverse strand [T₁₂MN] x M13 reverse primer

\[ T₁₂MN \]
\[
5' - (\text{T₁₂MN}) \quad \text{TTCAAGGTGA} \quad \text{AGTGGTTATG} \quad \text{ATTTATTTCC} \quad \text{AGTTTTAAC} \\
\text{CAGCTCAGTA} \quad \text{ATCTAGCTAA} \quad \text{TCCATGTTAC} \quad \text{TCAATGTTTTT} \quad \text{TCCACAGTGC} \\
\text{ACTAGAAGAA} \quad \text{CCCCCCCCC} \quad \text{CCCCCCCCC} \quad \text{CCCCCCCCC} \quad \text{CTGTGTTTAC} \\
\text{TACCGCGGGG} \quad \text{AGTGGTTATG} \quad \text{ATTTATTTCC} \quad \text{TCAATGTTTTT} \quad \text{TCCACAGTGC} \\
\text{GATT} \ldots \ldots \quad = 244 \text{bases}
\]

KEY: ACGT = internal primer sequence
ACGT = base present in manual sequence but not present in automated sequence
acgt = base present in automated sequence but not in manual sequence
discovered. However, attempts at manual sequencing using the internal primers gave poor results for both strands, particularly the forward strand. The problem may have been due to incomplete denaturation of the template during sequencing-primer annealing. Simple heat denaturation had been used up to this point, so it was decided to try alkaline denaturation using NaOH. Unfortunately, the results showed no improvement. It was possible that G/C rich regions within the sequence were preventing strand separation, or that a secondary structure was adversely affecting the ability of the template to take part in the sequencing reaction. A third alternative was that the sequence from which the primers were chosen was inaccurate. The generation of single-stranded template either by cycle sequencing or by cloning into M13 was deemed to be the next logical approach.

7.2.1.2 Automated cycle sequencing results

As described in Chapter 3, this technique produces a single-stranded template via linear amplification using a single primer for PCR. It was anticipated that this method would circumvent the problems caused by the use of double-stranded template that had occurred in earlier sequencing attempts. The Perkin Elmer ABI-PRISM Terminator™ Taq Primer cycle sequencing protocol was used in conjunction with T7 promoter and M13 reverse primers, in separate reactions. A positive control sample of M13 single-stranded DNA with M13 (-40) forward primer was also run with each batch of reactions. The test template readouts obtained produced very similar sequence data to that obtained from the manual sequencing reactions. See figure 7.1b. The internal reverse primer sequence was also preserved, with only a single base discrepancy. However, approximately 30% of the sequence from the internal forward primer showed variation between manual and automated sequencing results. This implied the previous failures with the internal forward primer could simply have been due to errors in primer sequence selection, but this still did not explain the poor results observed with the internal reverse primer. A new batch of the internal forward primer was made using the data obtained from the automated sequencing, and reactions were set up in duplicate using each of the internal primers. The resulting sequence data was poor, with a large number of unreadable bases present throughout each test file. The experiment was repeated using templates from pCRII/H8 plasmid, H8 PCR product (produced by amplifying across the pCRII insert-site using T7 promoter and M13 reverse primers) and isolated H8 insert (cut from pCRII with
NsiI). Each template was used in combination with T7 promoter, M13 reverse and the internal sequencing primers for each strand. The NsiI restriction fragment was only used with the internal primers as it did not possess annealing sites for the T7 promoter or M13 reverse primers. The results are shown in **table 7.1**, below.

**Table 7.1. Results of H8 template variation with cycle sequencing reaction**

<table>
<thead>
<tr>
<th>Template DNA</th>
<th>T7 promoter</th>
<th>M13 reverse</th>
<th>Internal primer forward strand</th>
<th>Internal primer reverse strand</th>
</tr>
</thead>
<tbody>
<tr>
<td>pCRII/H8</td>
<td>Good result</td>
<td>Good result</td>
<td>Reaction failed</td>
<td>Reaction failed</td>
</tr>
<tr>
<td>(0.5 µg)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pCRII/H8</td>
<td>Good result</td>
<td>Good result</td>
<td>Reaction failed</td>
<td>Reaction failed</td>
</tr>
<tr>
<td>(1.0 µg)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H8 PCR prod.</td>
<td>Poor result</td>
<td>Fair result</td>
<td>Reaction failed</td>
<td>Reaction failed</td>
</tr>
<tr>
<td>(300 ng)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H8/NsiI</td>
<td>Reaction failed</td>
<td>Reaction failed</td>
<td>[Not done]</td>
<td>[Not done]</td>
</tr>
<tr>
<td>(~ 0.5 µg)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[M13 positive control (0.4 µg) - Good result]

These results demonstrated that the internal primers could not be used in cycle sequencing with any of the H8 templates. Again, the forward strand was more problematic than the reverse strand. At this point it was decided to try cloning the H8 insert into M13 to generate single-stranded templates.

### 7.2.1.3 Sequencing using single-stranded M13 cloning products

A second method was applied to try to generate single-stranded DNA products for use in sequencing using M13 cloning. A full description of the principles and methods involved in M13 cloning are given in Chapter 3. The H8 sequence first had to be cut with appropriate restriction enzymes to facilitate its removal from the pCRII™ plasmid vector and subsequent
insertion into M13 mp18 RF vector. In order to achieve this, the G(C)G MAP program was applied to assemble a restriction map of the 473 base 3′ reverse strand sequence obtained from a] and b] in figure 7.1. Firstly, a single digest using EcoRI was used to cut out the entire H8 insert from pCRIITM, as both EcoRI sites lay within the pCRIITM flanking sequences, and produced a fragment of ~ 800 bases. Secondly, a double-digest using HindIII and XbaI was used to give the 5′ region of the H8 insert and produced a fragment of ~ 400 bases. The XbaI site was in 5′ pCRIITM flanking sequence while the HindIII site was ~ 400 bases upstream of the T12MN region. This latter combination also produced a second band of approximately the same size solely due to the action of HindIII. This second band represented the 3′ end of H8 and resulted from the presence of a second HindIII site in the 3′ flanking region of pCRIITM. See figure 7.2.
Chapter 7  Further analysis of the cDNA band H8

Initial attempts at cloning the restriction fragments into M13 were unsuccessful and no plaques containing recombinant clone were obtained. Steps were taken to optimise ligation of H8 insert to M13 vector. A range of ligation ratios of vector:insert from 1:3 to 1:10 were tried. Ligation was carried out using normal concentration (1 U/µL) and high concentration T4 DNA ligases (2000 U/µL). Ligation reactions were performed at different temperatures and for different durations, ranging from 3 hours at 20 °C to overnight at 16 °C. These parameter alterations made no discernible difference to the cloning outcome and still no clear plaques were forthcoming. It was then decided to investigate insert and vector preparation methods to determine if the ends of the molecules were somehow faulty and this was preventing ligation. The insert fragment ends may have been damaged as a result of excessive restriction enzyme activity or during gel purification. The vector ends may have similarly been over-eaten by the restriction enzymes or during dephosphorylation. The duration of all enzymatic reactions was kept to a minimum and various methods of purification of the insert fragment from agarose gel were employed. Exposure of the DNA samples to ethidium
bromide and to UV light was also minimised, as both of these factors can result in DNA damage. Unfortunately, there was still no improvement in the results. Changes were also made to the type of competent cells and lawn cells used, and also to the source of water. None of these had any positive effects. The ligation mix was then set up using crude restriction digests, omitting gel purification and vector dephosphorylation. This time numerous clear plaques were obtained.

PCR analysis of the double-stranded DNA from the bacterial pellets revealed 100% of the selected clones contained insert fragments of the correct size. See figure 7.3. Single-stranded DNA was then prepared and used in manual sequencing reactions with M13 (-40) forward primer, as described in the chapter 3. Sequence analysis revealed that the EcoRI fragment/M13 clone was from the forward (AP₁) strand, while the XbaI-HindIII/M13 and HindIII/M13 clones were from the reverse (T₁₂MN) strand. The orientation of the fragments and sequencing results are shown in figure 7.4. Sequences from 5' and 3' ends of the H8 insert were then aligned using G(C)G GELASSEMBLE and GELMERGE programs. This revealed an overlap between 5' and 3' sequence fragments of approximately 60 bases (see figure 7.5 a). The full sequence of the H8 cDNA band was finally resolved to give a 656 base fragment, excluding primer regions. See figure 7.5 b.

7.2.1.4 Database identification of H8 cDNA tag

Now armed with the full cDNA tag sequence, it was hoped that the database search would be more fruitful, but results still did not reveal the identity of H8. Again, the sequence matched only over short segments of ~20-50 bases. These matching segments did not even show any pattern that might have indicated the presence of some type of internal repeat. One interesting result was that almost all of H8 matched in segments of ~20-45 bases to a human DNA sequence from a P1 artificial chromosome (PAC) from the X-chromosome. The other target matches with the highest homologies were human DNA for presenilin 1, African swine fever virus and the human ABL gene. See table 7.2.

The segments of H8 that showed homology to PAC 107N3 were mostly within L1 repeat elements (types B1 and M4). The H8 sequence was put through RepeatMasker2 (http://ftp.genome.washington.edu/cgi-bin/RepeatMasker) which failed to find any repeat elements in the tag sequence. This may have been because the sequence tag was too short to
Agarose gel electrophoresis demonstrating the verification of recombinant M13/H8 insert by PCR.

DNA for PCR was extracted by boiling cell pellets, obtained from plaque culture in a host cell suspension, and amplifying some of this template with M13 primers. This was done to make sure the suspension contained the desired insert before proceeding any further with the isolation of single-stranded insert DNA. (See text).

It can clearly be seen here that all plaques selected did indeed contain an H8 insert of ~ 800 base pairs.
The bacterial pellets from each plaque examined contain recombinant H8/EcoRI fragment DNA.

**Key:**
- 
  - Kb = kilobase marker
  - Co = M13 positive control (no insert)
Figure 7.4. Manual sequence data from H8 restriction fragments cloned into M13 mp18 vector

a) H8/EcoR1 DNA x M13 (-40) primer = Forward strand (AP₁) sequence

\[5' - \text{AGCCAGCGAA TATGGGTTTT AACCTACTAC ACATCAATGT TTTTCCACA}
GTTGCCCCATA GAGAGCTGCT ATGGTTCACAC TCAAGTGAGG GAGAAAAAA AAAAGAAAGA
ATAATGGCAA GGAACCCACCA CCAACCAACCA TCCAATTACTG GACTCAACTA AGAGGAGGCTC
TACCTTTGAC TACCTTTATG TAGGAAAGTC TCTCTTATTA AAATCTCTTG GGCAAAAAA
TTCTACTGTT ATAAACAAATT TTTTATT \ldots \ldots \ldots 3'\]

b) H8/HindIII DNA x M13 (-40) primer = Reverse strand (T₁₂MN) sequence

\[5' - \text{TTCGCTGGCTAC AAAAAAT TTGTTATAAC AGTAGAAATT TTGTGCCCCAA GAGATTTTTA}
TAGAGAGAGAC TTACTCTAAAT TAAAGTCTAGG CTCCTCCTTAG TTGAGTCAGT
AATGGATGGC TGGGAGGGG GGGGTTGCTT GGGATCTTCTTTT TTTCTCTCTCT
CAGGAGATG CAGCCATAGTG ACGTCTCTTA GGAGCTGCTT GGAAAAAACA TTGATGTGTAG
TAGGTTAAA ACCCATA TTTTTTTTTTTTTGG \ldots \ldots \ldots 3'\]

c) H8/Xba-HindIII DNA x M13 (-40) primer = Reverse strand (T₁₂MN) sequence

\[5' - \text{TTTTTTTTTTTTGG CCTGTTTATG CCTAGCTTTA TTCAAGGTTA AGTGGTATGA}
TGTATTTCTCA AGGTATTTAC ACAGCTATCC ATGGGTTTAT TCAATACCC
TCAGAAAGTT CATTCCTTTTC TCTCTCTTAC CTCTCTGTTT TAATCAGTAT
GATATCCTTTA TATGGAAATA ACTATGAAAC TATCTCTCTT TATCTCTAAC
CTAGGACATA GCCGCTGCTG TTGAAAATAG TAGATATCAG TAGATATGTA
ATAGCTCTCA TCACCAGATT TCATTATAAT TTATATGCAT CAGGTTGTC
AATAACACAC TGGATATTTTC ATCTCTGTGGTCTCATACTCT GTATCATCTG
CGCTAAGAAGAAGTGGCTATGCACTAAT CACATATCTGCTA \ldots \ldots \ldots 3'\]

Key:
- AGCCAGCGAA = AP₁
- TTCGCTGGCT = AP₁ (reverse & complementary)
- TTTTTTTTTTTGG = T₁₂MN
- = H8 forward strand
- = H8 reverse strand
- = region of H8 sequenced
Figure 7.5. Complete H8 cDNA tag sequence

a) Consensus map o f sequenced fragments from G(C)G GELMERGE
program
7

6
5
4
3
2
C

+----------------------------------------

—----► = H8/HindIH DNA
» = H8/pCRII DNA

+
+
+

=H8/EcoRI DNA
= Consensus sequence

I

I

I

I

I

I

I

I

0

100

200

300

400

500

600

700

The consensus map shows where the sequenced fragments overlap.
The consensus line represents the entire sequence.

b) H8 cDNA tag sequence (forward & reverse)
[A P i ] 5' TATGGGTTTTAACCTACTACACATCAATGTTTTTCCCACAGTGCCCCTAAGAAGACTGCT

3-ATACCCAAAATTGGATGATGTGTAGTTACAAAAAGGGTGTCACGGGGATTCTTCTGACGA
ATGGTTCAACTCAAGTGAGAGGAGAAAAAAAAAAGAAAGAATATGCCAAAGGAACCACCA

TACCAAGTTGAGTTCACTCTCCTCTTTTTTTTTTCTTTCTTATACGGTTTCCTTGGTGGT
CCACCCACCATCCAATTACTGACTCAACTAAGAGGAGCTCTACTTTTGTCTACTTTATTG

GGTGGGTGGTAGGTTAATGACTGAGTTGATTCTCCTCGAGATGAAAACAGATGAAATAAC
TAGTAAAGTCTCTCTTATTAAAATCTCTTGGGCACAAAAATTCTACTGTTATAACAAATT

ATCATTTCAGAGAGAATAATTTTAGAGAACCCGTGTTTTTAAGATGACAATATTGTTTAA
TTTTATTAAAAAAGCTTAAACAGGAGCATAAAGATTTAGCAGATAATGTAGATAGCATTA

AAAATAATTTTTTCGAATTTGTCCTCGTATTTCTAAATCGTCTATTACATCTATCGTAAT
ACACTTTTCTTACAGCGCAGATGAACAGAGTATGAGACCAAGAGATGAAATACTCAGTGT

TGTGAAAAGAATGTCGCGTCTACTTGTCTCATACTCTGGTTCTCTACTTTATGAGTCACA
TTAGTTGCAACAACTGATGCAATAAATAATTAATGAAATCTGGTGATGAGAGATATTCAA

AATCAACGTTGTTGACTACGTTATTTATTAATTACTTTAGACCACTACTCTCTATAAGTT
ATCTTACTGATATCTACTTTATTCCAAGCAGCCTGCTATGTCCTAGGTTTAGGATAAAGA

TAGAATGACTATAGATGAAATAAGGTTCGTCGGACGATACAGGATCCAAATCCTATTTCT
AGATAAGTTCAATAGTTATTTCAATATAAAGTAATCATACTGATTAAAACAGGAAGGTAA

TCTATTCAAGTTATCAATAAAGTTATATTTCATTAGTATGACTAATTTTGTCCTTCCATT
ag aag aaaaag g aatg aatcttgg tag g tg atttg aataaacacatgg atag ctg tgg tt

TCTTCTTTTTCCTTACTTAGAACCATCCACTAAACTTATTTGTGTACCTATCGACACCAA
AAAACTTGGAATACAATCATAACACTTCACCTTGAATAAGCTAAGCATAAACCAGG3'

TTTTGAACCTTATGTTAGTATTGTGAAGTGGAACTTATTCGATTCGTATTTGGTCCs - [T 12MN]
BOLD

(A P j. STRAND)

NORMAL TYPE

= CODING STRAND

(T 12MN STRAND)

206

= NON-CODING STRAND


readily identify any repeats within it. A second reason could have been that the divergence between H8 and PAC sequences precluded H8 being recognised as a repeat element. The regions of PAC 107N3 that were matched to H8 neither contained an open reading frame nor appeared to be coding. This would imply that H8 lies within the 3' UTR. The PAC 107N3 sequence is from human chromosome Xq25. This is a Giemsa dark band and is very rich in repeats, the overall content of interspersed repeats being 67.3% [Personal communication from sequence depositor]. In view of this, it is not surprising that the matching regions were within some of these areas.

Table 7.2. Database sequence homologies of H8 cDNA tag

<table>
<thead>
<tr>
<th>Accession code:</th>
<th>Sequence identity:</th>
<th>Degree of homology:</th>
</tr>
</thead>
<tbody>
<tr>
<td>embZ75741HS107N3</td>
<td>Human DNA from PAC 107N3</td>
<td>64-81% over 20-45 base segments</td>
</tr>
<tr>
<td>dbjD84149D84149</td>
<td>Human DNA for presenilin 1</td>
<td>69% over one 44 base segment</td>
</tr>
<tr>
<td>gbU03762ASU03762</td>
<td>African swine fever virus</td>
<td>90% over one 30 base segment</td>
</tr>
<tr>
<td>gbU07562HSABLGR2</td>
<td>Human ABL gene</td>
<td>75-89% over 25-31 base segments</td>
</tr>
</tbody>
</table>

7.2.1.5 Peptide map of H8 sequence

The H8 nucleotide sequence was put through PEPDATA program in G(C)G to obtain possible amino acid sequences. The program allows the examination of all possible amino acid sequences within a nucleotide sequence where the coding region has not been identified, or where there is suspicion of frame-shift mutation within a known coding region. Results are shown in table 7.3.
Table 7.3  H8 amino acid sequence obtained using G(C)G PEPDATA program

This program allows the examination of all possible amino acid sequences within a nucleotide sequence where the coding region has not been identified, therefore

<table>
<thead>
<tr>
<th>Base</th>
<th>Amino Acid Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>IPKIG<em>CVT KRVSRGFF</em>R YQVEFTLLFF FSFLYGFLGG GGW<em>VND</em>VD</td>
</tr>
<tr>
<td>51</td>
<td>SPRDNR<em>NN IIISERILEN PCF</em>DDNIV* KIIFSNLSSY F*IVYYIYRN</td>
</tr>
<tr>
<td>101</td>
<td>CEKNSVASTCL ILWFSTL<em>VT NQRC</em>LRYLL ITLDHYSL* V <em>NDYR</em>NKVR</td>
</tr>
<tr>
<td>151</td>
<td>RTIQDPNPIS SIQVINKVIF H<em>YD</em>FCPSI SSFSLLRTIH *TYLCTYRHQ</td>
</tr>
<tr>
<td>201</td>
<td>F<em>TLC</em>YCEV ELIRFVG</td>
</tr>
<tr>
<td>251</td>
<td>YP KLDDV*LQKGGCHGSDDDTK LSSLSSFFFL</td>
</tr>
<tr>
<td>301</td>
<td>SYTVSLVVG GRLMTELILL EMKTDEITSF QRE<em>F</em>RTRV FKMTILFFK*</td>
</tr>
<tr>
<td>351</td>
<td>FFRICPRISK SSITSIVIVK RMSRLVSYL GSLLYESQIN VVDYVIY*LL</td>
</tr>
<tr>
<td>401</td>
<td>FTPLEPSTKL ICVIPDTNFE PYVISVWNL FDSYLV</td>
</tr>
<tr>
<td>451</td>
<td>TQNW MMCSYKKGVT</td>
</tr>
<tr>
<td>501</td>
<td>GILLTIPS<em>V HSPLFFFFFFFF RFPWMMVVVG **LS</em>FSSR* QKMK*HHFRE</td>
</tr>
<tr>
<td>551</td>
<td>YFMSHKSTIL TTTFINYFRP LLSISLE<em>L</em> MK*GSSDDTG SKSYFFYSSY</td>
</tr>
<tr>
<td>601</td>
<td>Q<em>SYISLV</em>L ILSFHFFFFFL T<em>NHPLNLFV YLSTPILNLM LVL</em>SOTYSI</td>
</tr>
<tr>
<td>651</td>
<td>RIWS</td>
</tr>
</tbody>
</table>

Key:  
- **aa** = amino acid  
- * = STOP codon  
- Other letters = standard amino acid single letter codes  

208
Further expression studies using H8 cDNA

A set of experiments was performed to determine the expression of the H8 mRNA in lung and in other tissues, both normal and malignant. These experiments included northern blotting and qualitative RT-PCR using H8 probe and primers, respectively. As the H8 cDNA had shown differential expression in pulmonary adenocarcinoma, it was decided to investigate whether it would do the same in other types of lung cancer and in non-pulmonary tumours.

RNA was isolated from matched samples of tissue from squamous cell, large cell and small cell carcinomas. RNA was also harvested from remaining tissue samples of the matched non-pulmonary tumours collected. These included adenocarcinomas from the colon, oesophagus, pancreas and stomach. Samples of RNA from normal liver and from ovarian adenocarcinoma were also included for analysis. The amounts of tissue left were small and the RNA yields from each were modest. Bearing in mind the RNA degradation problems described in Chapter 4, these samples were used immediately following purification. Fortunately, all the RNAs appeared to be completely intact when run out on a denaturing agarose gel. One aliquot of RNA (~ 10 µg) from each case was transferred to a nylon filter via northern blotting. Any remaining RNA (~ 1/20th of amount used in blot) was used for RT-PCR.

Northern blotting results using human tissue RNA samples

Hybridisation of H8 probe to RNA from other lung tumours produced a signal band of ~1.9 kb in size in each track. Initial observations indicated no apparent differential expression, as the signal from normal and tumour RNA lanes was of equivalent intensity. However, binding of the control probe GAPDH indicated marked differences in loading had occurred, with far less normal RNA being present in each sample than RNA from the tumour. The difference was apparent even with only visual inspection. This demonstrated that the H8 signal from the normal RNA lanes was proportionally greater than that from the tumour RNA for each type of lung tumour. See figure 7.6. Results from the non-lung tumour samples were slightly more mixed. Again, GAPDH indicated that less normal RNA was present in most of the samples, even though the hybridisation signal from the H8 probe was at least as intense in
Figure 7.6

Autoradiographs of a northern blot containing RNA extracted from matched human lung tumour tissues demonstrating hybridisation with probe H8. [Approximately 10 μg RNA loaded per lane.]

(a) Hybridisation signals from binding of the H8 probe indicate approximately equal expression levels of H8 in normal (N) and tumour (Ca) RNA samples for each type of tumour.

(b) The signal generated by hybridisation of the GAPDH control probe revealed a marked difference in the quantities of RNA loaded. Normal RNA was present in much lower amounts than tumour RNA. Consequently, if the loading irregularities are taken into consideration, then the relative level of H8 expression in all N samples must be greater than that in the Ca samples.

This experiment was performed to determine whether or not the differential expression of H8 was confined only to pulmonary adenocarcinoma.
the normal RNA lanes as it was in the tumour lanes. See figure 7.7. Results are summarised in table 7.4.

Table 7.4. Summary of northern blotting data using human tissue.

<table>
<thead>
<tr>
<th>RNA source</th>
<th>GAPDH probe signal</th>
<th>H8 probe signal</th>
<th>Relative intensity of H8 signal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Ca</td>
<td>N</td>
</tr>
<tr>
<td><strong>Lung samples:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCC</td>
<td>+</td>
<td>+++</td>
<td>++</td>
</tr>
<tr>
<td>LCC</td>
<td>+</td>
<td>+++</td>
<td>++</td>
</tr>
<tr>
<td>SCLC</td>
<td>+</td>
<td>+++</td>
<td>++</td>
</tr>
<tr>
<td><strong>Non-lung samples:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colon AC</td>
<td>+</td>
<td>++</td>
<td>+++</td>
</tr>
<tr>
<td>Oesophagus AC</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
</tr>
<tr>
<td>Pancreas AC</td>
<td>+/-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Stomach AC</td>
<td>+</td>
<td>++</td>
<td>+++</td>
</tr>
<tr>
<td>Liver (N)</td>
<td>+</td>
<td>(x)</td>
<td>+++</td>
</tr>
<tr>
<td>Ovary AC (Ca)</td>
<td>(x)</td>
<td>+++</td>
<td>(x)</td>
</tr>
</tbody>
</table>

Key: N = normal RNA sample  
Ca = tumour RNA sample  
AC = adenocarcinoma  
x = not done (sample not present)  
n/a = not applicable
Figure 7.6 Northern blot hybridisations with matched human lung tumour RNA samples.

a) H8 probe

- SCC = Squamous cell carcinoma RNA
- LCC = Large cell carcinoma RNA
- SCLC = Small cell carcinoma RNA
- N = normal lung RNA
- Ca = lung tumour RNA

~1.9 kb

b) GAPDH probe

Key:

SCC = Squamous cell carcinoma RNA  N = normal lung RNA
LCC = Large cell carcinoma RNA  Ca = lung tumour RNA
SCLC = Small cell carcinoma RNA
Figure 7.7

Sections from an autoradiograph of a northern blots containing RNA from various matched human tumour tissues demonstrating hybridisation with probe H8. The tumours from the oesophagus, stomach, pancreas, ovary and colon were adenocarcinomas. The lung tumour was a squamous cell carcinoma and the liver RNA was taken from a normal organ. [Approximately 10 μg RNA loaded per lane.]

As in figure 7.6, when the loading inequalities manifested by the variation in GAPDH signal are taken into consideration, a number of these non-pulmonary tumours also showed differential expression of H8. This was seen quite clearly in the matched normal RNA samples from the stomach, lung and colon, also in the normal liver sample. This demonstrated that the expression difference is not just confined to pulmonary tissues.

A second band (~ 0.5 kb) is present in several of the lanes. This may represent a splice variant of the H8 mRNA.
Figure 7.7 Human tissue RNA samples hybridising with H8 cDNA probe.

<table>
<thead>
<tr>
<th>H8 probe</th>
<th>Oesophagus</th>
<th>Stomach</th>
<th>Lung(SCC)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N Ca</td>
<td>N Ca</td>
<td>N Ca</td>
</tr>
<tr>
<td>H8 ~ 1.9 kb</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>?splice variant</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

GAPDH probe (for same samples)

<table>
<thead>
<tr>
<th></th>
<th>Colon</th>
<th>Liver</th>
<th>Pancreas</th>
<th>Ovary</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N Ca</td>
<td>N only</td>
<td>N Ca</td>
<td>Ca only</td>
</tr>
<tr>
<td>GAPDH ~ 1.2 kb</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

GAPDH probe (for same samples)
7.2.2.2 Qualitative RT-PCR using primers designed from H8 sequence data

The H8 sequence data was used to design a set of primers for RT-PCR using the G(C)G PRIME program. These consisted of a single forward primer (F1) and two reverse primers (R1 & R2), their positions and sequence information are shown in figure 7.8. Primer R2 was specifically designed for use in 5' RACE-PCR, see below. The F1/R1 amplicon was 440 bases and the amplicon for F1/R2 was 134 bases in length. The primer pairs were first used with templates from cell line L-132 genomic DNA and H8/PCRII plasmid DNA. These samples acted as positive controls to ensure the primers amplified the target sequences correctly. Next, the primers were tested with poly(A) RNA from L-132 and SKLU-1 using solid-phase RT-PCR from magnetic beads. See Chapter 3, section 3.10. This was done as a further confirmatory test to ensure that H8 was a cDNA and not a genomic contaminant. Finally, RT-PCR was performed using cDNAs made from some of the tissue samples listed above. RNA extracted from breast cell lines HBL-100 and MDA-468 and colonic cell line HCEC were also included in this experiment.

The results are shown in figure 7.9. Amplification of the positive control DNA from L-132 and H8/plasmid yielded PCR products of the correct size with each primer combination. Poly(A) RNA from both cell lines also produced a single band of the correct size with each primer pair. The tissue cDNAs were only amplified with F1/R1 and also produced single discrete bands of the expected size. Some of the cDNAs did show mild levels of genomic DNA contamination as revealed by PCR product in the RT negative controls. PCR product from normal and tumour cDNAs gave bands of similar intensities, but this in no way represents the initial quantities of H8 mRNA in the sample. Any target sequences present within the sample will be amplified exponentially during this type of PCR, so the signal intensity of each PCR product formed is not proportional to the original amount. This reaction was intended to be purely qualitative and not quantitative.

7.2.3 5' RACE-PCR

In order to try and determine the full length H8 cDNA sequence 5' RACE-PCR was attempted. A gene specific internal primer (R2) was designed to amplify the sequence at the 5' end of the cDNA. See figure 7.8. This method is widely used to determine additional sequence information and to identify full cDNA sequence without the need for cDNA library
Figure 7.8  PCR primers designed from H8 sequence

Key:  
- RT-PCR primer  
5' RACE PCR primer
Figure 7.9 (a) and (b)

Agarose gel electrophoresis demonstrating the results of reverse transcription PCR using primers made from the H8 sequence on various human templates. Two primer combinations, \( F_1/R_1 \) & \( F_2/R_2 \), were used to amplify 440 bp and 134 bp regions of the target, respectively.

(a) First, a positive control reaction was set up using genomic DNA from cell line L-132 (normal) and the plasmid containing the H8 insert. PCR products of the correct size were amplified from each template. This verified that the primers worked.

(b) Secondly, the reactions were repeated using mRNA template from cell lines L-132 (normal) and SKLU-1 (tumour). Again, all the reactions were successful. PCR product bands of equal intensity are visible in both normal and tumour template lanes. This reflects the fact that this type of PCR is purely qualitative. Quantitative assays using different protocols could be performed to determine the relative amounts of product in each sample.

This result further verifies H8 as a cDNA originating from an mRNA transcript.
Figure 7.9  Results of RT-PCR using H8 primers

a) Positive control PCR using genomic L-132 & H8/pCRII plasmid DNA samples

Lanes 1-3 = genomic L-132 DNA, 4-6 = H8/pCRII plasmid DNA
Lanes 1& 4 = Ft/R2 (134 bp product), 2& 5 = Ft/R1 (440 bp product)
Lanes 3& 6 = negative controls (no DNA template)

b) Solid phase RT-PCR using poly(A) RNA from L-132 & SKLU-1 cell lines

Lanes 1-3 & 7-8 = L-132 cDNA, lanes 4-6 & 9-10 = SKLU-1 cDNA
Lanes 1/ 4 = Ft/R2, lanes 2/ 5 = Ft/R1
Lanes 3/ 6 = PCR positive control using K-ras primers (~300 bp product)
Lanes 7/ 9 & 8/ 10 = negative controls for Ft/R2 & Ft/R1 respectively
Figure 7.9 (c)

Finally, the reactions were repeated again using various human templates from tissues and cell lines. Here, the negative controls (odd numbers) are run in lanes adjacent to the test samples (even numbers). Some of the samples bands in the negative control lanes, indicative of probable genomic DNA contamination. The vast majority, however, do not. As in (b), PCR product bands of equal intensity are visible in both normal and tumour template lanes.

This experiment helped to verify the results observed in the northern blots of figure 7.7, but quantitative RT-PCR would have to be performed to prove that differential expression of H8 was present.
c) RT-PCR using matched human tissue cDNA templates

Lane 1& 2 = Liver (N), 3& 4 = Ovary (Ca)
Lane 5& 6 = Squamous cell cancer lung [SCC] (N), 7& 8 = SCC (Ca)
Lane 9& 10 = Oesophagus (N), 11& 12 = Oesophagus (Ca)
Lane 13& 14 = Stomach (N), 15& 16 = Stomach (Ca)
Lane 17& 18 = Small cell cancer lung [SCLC] (N), 19& 20 = SCLC (Ca)
Lane 21& 22 = Breast cell line (N), 23& 24 = Breast cell line (Ca)
Lane 25& 26 = Colon cell line (N)
Odd numbered lanes = RT negative controls (no M-MLV in cDNA step)
N = normal matched tissue  Ca = tumour tissue
screening. Amplification of the 5' region of the cDNA tag was felt to be particularly important as this end would contain information on any promoter sites, open reading frames, etc. It was anticipated that it would also result in more informative sequence identification following database scans than had been obtained previously from using the 3' sequence.

Initially, a fresh sample of normal human liver poly(A) RNA was used as the template. The RNA had been cleaned before use to remove any traces of genomic DNA. However, attempts at amplification did not proceed easily. This can occur with sequences possessing “difficult” 5' ends. Random hexamers were substituted and first strand cDNA synthesis was repeated using normal breast RNA, as successful amplification of this template with H8 RT-PCR primers had been achieved previously. Unfortunately, this made no difference to the result. The problem was either in the ligation step or in the PCR. Attention was turned first to the ligation step. This was repeated with the original liver cDNA and this time the ligation reaction was left at room temperature (25°C) for 3 hours. Subsequent attempts at PCR with this sample remained unsuccessful. Regrettably, no further laboratory investigation of the reaction failure could be undertaken as the project had approached its termination deadline.

### 7.2.4 In situ hybridisation

A preliminary study to determine the viability of the lung sections for use with in situ hybridisation (ISH) was performed using a mitochondrial-RNA specific probe. A proteinase K (PK) titration was also undertaken, and 1, 2, 5 or 10 μg/ml PK solution were added to the sections. Lung sections from the urethane-treated mice were used for this experiment. The mouse lung tissues had been dissected out rapidly and immediately fixed in 4% paraformaldehyde to maximise the survival of mRNA within the tissue. (See Chapter 3, section 3.1.3.) Hybridisation of the mitochondrial probe was reasonably successful, and PK levels of 1 and 2 μg/ml demonstrated optimal digestion. An oligo cocktail containing seven H8 sequences was then designed according to the method described in Chapter 3, section 3.12.1. See table 7.5 for details.
Table 7.5 Anti-sense H8 oligo sequences used for in situ hybridisation.

<table>
<thead>
<tr>
<th>Oligo:</th>
<th>Sequence: (5'-3')</th>
<th>(T_m) °C:</th>
<th>GC %:</th>
</tr>
</thead>
<tbody>
<tr>
<td>393</td>
<td>TTA ATT ATT TAT TGC ATC AGT TGT TGC AAC</td>
<td>63.3</td>
<td>27</td>
</tr>
<tr>
<td>423</td>
<td>GAT TTG AAT ATC TCT CAT CAC CAG ATT TCA</td>
<td>64.6</td>
<td>33</td>
</tr>
<tr>
<td>453</td>
<td>GGC TGC TTG GAA TAA AGT AGA TAT CAG TAA</td>
<td>63.1</td>
<td>37</td>
</tr>
<tr>
<td>483</td>
<td>TCT TCT TTA TCC TAA ACC TAG GAC ATA GCA</td>
<td>62.9</td>
<td>37</td>
</tr>
<tr>
<td>595</td>
<td>CAG CTA TCC ATG TGT TTA TTC AAA TCA CCT</td>
<td>66.0</td>
<td>37</td>
</tr>
<tr>
<td>625</td>
<td>GTG TTA TGA TTG TAT TCC AAG TTT TAA CCA</td>
<td>62.5</td>
<td>30</td>
</tr>
<tr>
<td>655</td>
<td>CTG GTT TAT GCT TAG CTT ATT CAA GGT GAA</td>
<td>65.4</td>
<td>37</td>
</tr>
</tbody>
</table>

Key:

- \(T_m\) = melt point temperature
- GC = proportion of [guanine + cytosine] residues in sequence

As laboratory time was rapidly approaching its conclusion, sections of lung from human tumour cases were used directly with this H8 probe cocktail. PK levels of 1 and 2 µg/ml and H8 probe at a concentration of 500 ng/ml was used. Human tonsil hybridised with mitochondrial probe was used as a control for the reaction, and human lung hybridised with mitochondrial probe was used as a control for the lung tissue. While the hybridisation of mitochondrial probe to human tonsil gave a positive result, the result using human lung was less satisfactory. Apart from occasional signal in some white blood cells, no hybridisation was observed. Results from human lung with the H8 cocktail were similarly negative. As with the RACE-PCR, the imminent cessation of the project regrettably precluded any further analysis of this experimental system.
7.3 Discussion

What exactly was the H8 sequence? The H8 cDNA tag was identified in apparently normal human lung tissue. Normal refers to lung tissue that did not contain any overt signs of neoplastic change, either on macroscopic or microscopic examination. The process whereby this cDNA tag was produced, namely DDRT-PCR, is prone to a high rate of false-positives, as well as to contamination. In order to verify that H8 was indeed a cDNA, the following points must be considered. Firstly, the RNA from which H8 was isolated was treated with DNase as part of its purification and the samples had OD\textsubscript{260/280} values of 1.8-1.9, which implied that they were of a high quality and free from DNA contamination. Secondly, DDRT-PCR is designed to select for the poly(A) tails of mRNAs via the Ti\textsubscript{12}MN anchor primer, although, it must be remembered that annealing to any sequence containing internal runs of A bases cannot be ruled out. Thirdly, control RT-PCR with K-ras primers produced only one discrete band, appropriate for the amplification of a cDNA template. (See Chapter 4, figure 4.9.) All these factors would indicate that H8 did not contain any genomic contaminants, at least prior to the DDRT-PCR step.

The results obtained from the cell line total RNA northern blot showed a three-fold greater expression of H8 in the normal RNA relative to the tumour RNA, as measured by densitometry. (See Chapter 6, figure 6.5a.) The results from the non-adenocarcinoma northern blots gave similar results. (See figure 7.6.) Let us consider the data from the northern blots further. If the H8 probe was binding to genomic DNA within the RNA transferred onto the nylon membrane, it would have appeared as a smeared signal with a high molecular weight. H8 gave a clear and relatively clean signal of approximately 1.9 kb size. If the H8 probe was binding to ribosomal RNA within the blot sample it would have produced a highly intense signal. The signal observed from H8 probe hybridisation was obvious and well defined but was not excessively strong, becoming visible after autoradiographic exposures of 24-48 hours. One would also have expected any homology to ribosomal RNA to have been picked up on the database searches and this simply did not occur. Also, the hybridisation band produced from the Southern blot (2.1 kb) was larger than that produced from the northern blot (1.9 kb). This would imply that the genomic sequence contained regions which were lost in the mRNA transcript, i.e. introns.
Chapter 7  Further analysis of the cDNA band H8

The definitive experiment would have been to use the H8 probe on a poly(A) RNA northern blot membrane. This was tried, but technical difficulties in sample preparation prevented the isolation of mRNA of suitable quality in quantities sufficient for a northern blot. It was for this reason that solid-phase RT-PCR from magnetic beads was performed. The beads were used to capture poly(A) RNA and cDNA synthesis was carried out with the mRNAs still anchored to the beads. The mRNA was then removed, leaving only cDNA for use in RT-PCR. Numerous washing steps were performed to remove any genomic DNA contaminants. This method did not provide the same information as a poly(A) northern blot, but it was a perfectly valid alternative. RT-PCR using this approach with H8 primers clearly demonstrated the successful amplification of cDNA produced from poly(A) RNA. (See figure 7.9b.) If time had permitted, quantitative competitive RT-PCR could also have been attempted to determine the relative amounts of H8 PCR product from normal and tumour samples. Hopefully, this would have reinforced the expression differences revealed by the northern blotting experiments.

The fact that H8 did not share homologies with any of the database sequences is also not unusual. It is well known that DDRT-PCR preferentially amplifies short cDNA tags from the 3' end of mRNAs. This is only to be expected since the T12MN anchor primer is selecting mRNAs via the poly(A) tail. A large number of these 3' tags lie in the untranslated region (UTR) of the mRNA. As described in Chapter 2, this region is not always included in database sequences and shows great variation, even between homologous genes. Also, the databases are far from complete. Therefore, the absence of a positive sequence match does not give any reason to suppose that H8 is anything other that a cDNA tag from such a 3' UTR. A similar argument can be applied to explain why H8 does not have an open reading frame (ORF). It was hoped that 5'RACE-PCR would help determine more information about the 5' end of the cDNA, and even allow the complete sequence to be elucidated. Unfortunately, there was insufficient time to pursue this.

Some sequences are inherently difficult to clone and sequence, and H8 was one of these. This was demonstrated in the attempts at manual and automated sequencing. The reverse strand (T12MN/ 3' end) always provided more sequencing data than the forward strand.
(AP1/5' end). The forward strand was difficult to sequence beyond the region ~120-150 bases downstream of the 5' end. Secondary structures within this region of the molecule may have inhibited the formation of sequencing products. The problems encountered with the preliminary 5' RACE-PCR could also be explained by these factors. The presence of secondary structures or a high G/C content can make RACE, particularly the 5' reaction, extremely difficult to accomplish. Whether it was the 5' end of the cDNA tag or the 5' end of the full length cDNA that was causing the problem here cannot be known. The difficulties observed with single-stranded cloning of H8 are less clear. Again, some sequences simply do not lend themselves to being manipulated in this manner for seemingly idiosyncratic reasons, and do not form stable ligation products. It is interesting to note that the cloning only succeeded when crude restriction products and very approximate ligation ratios were used.

The ISH results using the H8 cocktail on human lung were most likely due to degradation of the mRNA targets that had occurred before the sections were prepared. This was reflected by the problems encountered in Chapter 4. Alternatively, the H8 target mRNA may have been relatively short-lived and have degraded by the time the tissues were fixed and sectioned. Had time allowed, more use could have been made of the mouse sections, as well as the large number of unmatched human pulmonary adenocarcinoma tissues collected. In addition, frozen sections could have been obtained and ISH could have been performed using cytospin samples from the various lung cell lines. Riboprobes could have been used instead of an oligo cocktail.

In summary, here is what is known about cDNA tag H8;

i) It was identified in normal RNA from human pulmonary adenocarcinoma via DDRT-PCR.

ii) It is a cDNA sequence derived from an mRNA of ~1.9 kb in size.

iii) The cDNA tag is probably derived from a 3'UTR, as indicated by lack of homology to known sequences and the absence of an ORF.

iv) The genomic DNA appears to be ~ 2.1 kb.

v) The mRNA shows diminished expression levels in adenocarcinomas from lung, stomach, and colon, and also in other types of lung tumours, when compared to the normal (non-tumour) tissues and may represent a tumour suppressor gene.
7.4 General discussion

This project has shown that DDRT-PCR can be used to study gene expression in human lung cancer. The results obtained from human tissue case #1 demonstrate what can be achieved with a good quality template. More such gels could have been produced were it not for the setbacks encountered during the preparation of some of the other human tissue RNA samples. Tremendous cooperation was given by all the clinical staff during the tissue collection process, but one still has to wait for suitable cases to be operated upon. Acquisition of such cases is an unpredictable variable and can take many months. If more time had been available, perhaps fresh starting material could have been collected. It was for this reason that the role of the cell lines was extended. They gave excellent DDRT-PCR results and provided the first evidence that the experimental methods were sound. The use of lung epithelial cell line L-132 as a normal (i.e. non-tumour) control was not perfect. It could be argued that any immortalised cell line is definitely not normal and it may not arise from the same cell of origin as the tumour line. The use of L-132 was, however, considered to be the only feasible option at the time. It could also be argued that any cell cultures used should have been obtained from the same patient, as cultures originating from different people will show a certain amount of genetic variation. However, this may be quite difficult to achieve in practice. Behaviour in vitro will not be the same as that in vivo for any cell type, and how representative of the tissue of origin is any established cell line?

Several factors, both intrinsic and extrinsic, prevented some of the experiments from running entirely as anticipated, but even this had its positive side. It is precisely when things do not go according to plan that one is forced to examine every aspect of the experimental design. This creates a better understanding of the processes involved and can even result in the creation of a superior system.

If the project could be done again, what should be done differently? The use of different template sources, such as primary cultures or the micro-dissection of specific cell populations, is the first point. This would simplify the differential display comparisons and might help to maximise positive results. Incorporating some of the modifications to the
DDRT-PCR protocol, or even using one of the newer techniques described in Chapter 2 could also reduce time spent analysing false positives.

If more time was available to continue the work on H8, several additional experimental strategies could be employed to learn more about the sequence. The elucidation of the full length cDNA would be the first objective. This could be done by RACE-PCR combined with cloning and sequencing, or by more conventional cDNA library screening. Database homology could then be re-checked for matches to known sequences which may provide further information as to the function of the H8 product. If the cDNA contained an open reading frame for an amino acid chain, analysis of the protein structure could provide more clues to its function. Investigation of in vitro translation could also be useful to determine whether the product was functional. The optimisation of the in situ hybridisation method to localise of the cell type expressing the H8 mRNA would be an obvious experiment to perform. Chromosomal localisation of the H8 gene could then be achieved by the use of somatic cell hybrids, while fluorescent in situ hybridisation could be performed on metaphase plates. Transfection of sense and anti-sense constructs into tumour cell cultures could provide evidence of any tumour suppressor activity.

7.5 Conclusions

Differential display RT-PCR is not a perfect technique by any means. At the start of this project, DDRT-PCR was still a new and relatively untested method and no-one was really aware of the pitfalls. Going into unknown territory was what was so exciting about using a state-of-the-art method. However, DDRT-PCR does provide the best opportunity to discover novel genes in a given system in a relatively short time. It is also particularly useful when template sources are limited. In addition, it incorporates many different molecular biology techniques, from the basics of nucleotide extraction to the more sophisticated procedures of RACE-PCR and in situ hybridisation. For these reasons, DDRT-PCR is ideal in providing a comprehensive training experience in core molecular biology techniques. In terms of man-hours spent working up the techniques and in the numbers of false-positives observed, the results obtained here have been entirely consistent with those from other studies using DDRT-PCR.
The controversies in lung cancer biology will continue. Every year the lung cancer incidence figures grow larger, every month a new lung cancer gene is cited in the literature, and almost every week we learn a little more about this group of tumours. Techniques like DDRT-PCR will, no doubt, continue to make significant contributions to this core of knowledge. Eventually, all the individual pieces of information that are uncovered by groups worldwide will reach a critical mass and, rather like a jigsaw puzzle, will fall into place to reveal the whole story of lung tumourigenesis. It is only a matter of time.
References


Bauer D. Identification of differentially expressed mRNA species by an improved display technique (DDRT-PCR). *Nucleic Acids Research* 1993; 21 (18): 4272-4280

Beer D.G, Malkinson A.M. Genetic influence on type 2 or Clara cell origin of pulmonary adenomas in urerthan-treated mice. *Journal of the National Cancer Institute* 1985; 75: 963-969


Bepler G, Garcia-Blancer M.A. Three tumour suppressor regions on chromosome 11p identified by high resolution deletion mapping in human NSCLC. *Proceedings of the National Academy of Science USA* 1994; 91: 5513-5517

Birnboim H.C, Dloy J. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research* 1979; 7: 1513


References


228

Croce C. Genetic approaches to the study of the molecular basis of human cancer. *Cancer Research* 1991; 51(s): 5015s-5018s


Debouck C. Differential display or differential dismay? *Current Opinions in Biotechnology* 1995; 6: 597-599


Faraldo M.J, Dux A, Muhlbock O, Hart G. Histocompatibility genes (H-2 complex) and susceptibility to spontaneous lung tumors in mice. *Immunogenetics* 1979; 9: 383-404


FFrench-Constant C. Alternative splicing of fibronectin- many different proteins but few different functions. *Experimental Cell Research* 1995; 221: 261-271


References


Hecht S.S. Chemoprevention by isothiocyanates. *Journal of Cellular Biochemistry* 1995; 22(s): 195s-209s


References


References


Kwaik Y.A, Pederson L.L. The use of differential display-PCR to isolate and characterise a *Legionella pneumophila* locus induced during the intracellular infection of macrophages. *Molecular Microbiology* 1996; 21(3): 543-556

LaForgia S. Receptor protein tyrosine phosphatase G is a candidate tumor suppressor gene at human chromosome region 3p21. *Proceedings of the National Academy of Science USA* 1991; 88: 5036-5040


Li F, Barnathan E, Kariko K. Rapid method for screening and cloning cDNAs generated in differential mRNA display: application of northern blot for affinity capturing of cDNAs. *Nucleic Acids Research* 1994b; 22 (9): 1764-1765


Malkinson A.M, Nesbitt M.N, Skamene E. Susceptibility to urethan-induced pulmonary adenomas between A/J and C57BL/6J mice: use of AXB and BXA recombinant inbred lines indicating a three-locus genetic model. *Journal of the National Cancer Institute* 1985; 75: 971-974


Mardon H, Grant R, Grant K, Harris H. Fibronectin splice variants are differentially incorporated into the extracellular matrix of tumorigenic and non-tumorigenic hybrids between normal fibroblasts and sarcoma cells. *Journal of Cell Science* 1993; 104: 783-792
References


Merlo A. Frequent microsatellite instability in primary SCLC. Cancer Research 1994; 54: 2098-2101


Mistudomi T, Steinberg S.M, Nau M.M. p53 gene mutations in NSCLC cell lines and their correlation with the presence of ras mutations and clinical features. Oncogene 1992; 7: 171-80


Mulshine J.L. Initiators and promoters of lung cancer. *Chest* 1993; 103: 4s-11s


References


Reissmann P.T. Inactivaiton of the *Rb* susceptibility gene in NSCLC. *Oncogene*. 1993; 1913-1919


References


Sager R. Expression genetics in cancer: shifting the focus from DNA to RNA. *Proceedings of the National Academy of Science USA* 1997; 94: 952-955


References


Sawaya M.E. Biochemical mechanisms regulating human hair growth. *Skin Pharmacology* 1994; 7: 5-7


Shuguang L. Analysis of polyaromatic hydrocarbons in cooking oil fumes. *Archives of Environmental Health* 1994; 49(2): 119-122


Southern E.M. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology 1975; 98: 503-517*


Strong L.C. The establishment of the "A" strain of inbred mice. *Journal of Heredity 1936; 27: 21-24*
References


Suda Y, Aizawa S, Hirai S. Driven by the same Ig enhancer and SV-40 T promoter: *ras* induced lung adenomatous tumours, myc-induced pre-B-cell lymphomas and SV40 large T gene a variety of tumours in transgenic mice. *EMBO Journal* 1987; 6: 4055-4065


Takahashi T. The *p53* gene is very frequently mutated in SCLC with a distinct nucleotide substitution pattern. *Oncogene.* 1991; 6: 1775-1778


References


249


Wattenberg L.W. Inhibition of chemical carcinogen-induced pulmonary neoplasia by butylated hydroxyanisole. *Journal of the National Cancer Institute* 1973; 50: 1541-1544


Wong K.K, McClelland M. Stress-inducible gene of \textit{Salmonella typhimurium} identified by arbitrarily primed PCR of RNA. *Proceedings of the National Academy of Science USA* 1994; 91: 639-643


Appendix I - Materials

Amplitaq DNA polymerase
Antidigoxigenin antibody
AR tips
Autorad RX film
\(\alpha^{35}\) S-dATP
\(\alpha^{32}\) P-dCTP
Calf intestinal alkaline phosphatase
Denaturing gel mix
Digoxigenin-11-dUTP
Discovery Northern Territory membrane
DNaseI (RNase-free)
dNTPs (100 mM stock)
DTT
Dynabeads
Gel-Bond PAG™ film
Hexamers (random)
IPTG
JM105 bacterial cells
Klenow fragment
M13 mp18 RF DNA
3 MM paper
Marathon™ cDNA kit
Mineral oil
M9 minimal salts
M-MLV enzyme & 5x buffer
Multiprime kit
Non-denaturing gel mix
Nytran® membrane
One-Phor-All 10x buffer
PCR buffer

Perkin Elmer
Boehringer
NBL
Fuji
Amersham
Amersham
Gibco
Scotlab
Boehringer
Invitrogen
Biogene
Pharmacia
Gibco
Dynal
FMC
Pharmacia
Biogene
Pharmacia
Sigma
Whatman
Clontech
Sigma
Sigma
Gibco
Amersham
Biorad
Anderman
Pharmacia
Perkin Elmer
<table>
<thead>
<tr>
<th>Item</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>QIAEX II kit</td>
<td>Qiagen</td>
</tr>
<tr>
<td>Restriction enzymes <em>EcoRI/HindIII</em></td>
<td>Pharmacia</td>
</tr>
<tr>
<td>Restriction enzymes <em>NsiI/SmaI/BamHI</em></td>
<td>Boehringer-Mannheim</td>
</tr>
<tr>
<td>RNA map kit™ A</td>
<td>Biogene</td>
</tr>
<tr>
<td>RNASin</td>
<td>Promega</td>
</tr>
<tr>
<td>RNAzol™ B</td>
<td>Biogene</td>
</tr>
<tr>
<td>Sequenase™ 2.0 kit</td>
<td>Amersham</td>
</tr>
<tr>
<td>T4 DNA ligase &amp; 5x buffer</td>
<td>Gibco</td>
</tr>
<tr>
<td>TA cloning™ kit</td>
<td>Invitrogen</td>
</tr>
<tr>
<td>Terminal deoxytransferase (recombinant)</td>
<td>Gibco</td>
</tr>
<tr>
<td>Thin-walled PCR tubes (0.5 ml)</td>
<td>Alpha Labs</td>
</tr>
<tr>
<td>Tracker™ tape</td>
<td>Amersham</td>
</tr>
<tr>
<td>Virkon®</td>
<td>Day-Implex</td>
</tr>
<tr>
<td>X-gal</td>
<td>Biogene</td>
</tr>
</tbody>
</table>

All other chemicals were obtained from Sigma or Fisher.

Cell lines L-132 and SKLU-1 were purchased from European Collection of Animal Cell Cultures, Centre for Applied Microbiology and Research.
Appendix II - Primer sequences

K-ras primers for RT-PCR control

#1 (exon 1) 5' TAA TTG TAA GGC CTG CTG AAA ATG ACT G A G 3'
#2 (exon 2) 5' ATA CAC AAA GAA AGC CCT CC 3'

Product = 286 bp length (with cDNA template).

Arbitrary primers (APs) for DDRT-PCR

AP1 5' AGCCAGCGAA 3'
AP2 5' GACCGCTTGT 3'
AP3 5' AGGTGACCGT 3'
AP4 5' GGTACTCCAC 3'
AP5 5' GTTGCGATCC 3'

Sequences were obtained from Bio/Gene Ltd, UK (distributor for GenHunter, USA).

Sequencing primers

M13 (-40) Forward 5' GTTTTCCCAGTCACGAC 3'
M13 Reverse 5' CAGGAAACAGCTATGAC 3'
T7 Promoter 5' TAATACGACTCACTATAGGG 3'
Appendix III - Buffers, media and other reagents

1. Source of enzyme reaction buffers

5x Labelling buffer (Amersham)
5x Ligation buffer (Invitrogen/ Gibco)
10x One-Phor-All buffer* (Pharmacia)
10x PCR buffer (Perkin Elmer)
5x RT buffer (Gibco)

All buffers were aliquoted and stored at -20°C (in a non-frost free freezer).
* Stored at 4°C.

2. Composition of media & plates for cloning

L-broth (Luria-Bertani medium) per litre:

10g Bacto®-tryptone
5g Bacto®-yeast extract
5g NaCl

The solution was adjusted to pH 7.5 with NaOH and autoclaved. [+/− ampicillin 100 μg/ml once solution has cooled to 50°C if required.] Solution stored at room temperature.

L-agar plates (LB plates) per litre:

Add 15g agar to 1L of L-broth.

The solution was adjusted to pH 7.5 with NaOH and autoclaved. [+/− ampicillin 100 μg/ml once solution has cooled to 50°C if required.] It was then aliquoted into 90 mm petri dishes and allowed to set at room temperature. Plates covered in saran wrap and stored at room temperature for ~1 week (or at 4°C for up to one month). Plates should be left in 37°C incubator for 1 hour to dry before use.
M9 minimal salts per litre:

- 6.0 g \( \text{Na}_2\text{HPO}_4 \)
- 3.0g \( \text{KH}_2\text{PO}_4 \)
- 0.5g \( \text{NaCl} \)
- 1.0g \( \text{NH}_4\text{Cl} \)
- 15g agarose

Sterile \( \text{H}_2\text{O} \) to ~1L was added and the solution was autoclaved.

M9 minimal medium per litre:

To 1L of M9 minimal salts the following filter-sterilised components were added;

- 2.0 ml 1M \( \text{MgSO}_4 \)
- 0.1 ml 1M \( \text{CaCl}_2 \)
- 10.0 ml 20% glucose
- 1.0 ml 1M Thiamine-HCl

The medium was then stored at 4°C.

3. Reagent recipes

Denaturing buffer for RNA:

The following reagents were mixed together in a fume hood using an RNase-free tube;

- 1000 \( \mu\text{l} \) De-ionised formamide
- 350 \( \mu\text{l} \) 37% formaldehyde
- 200 \( \mu\text{l} \) 10x MOPS buffer

This was then aliquoted x 50 -100 \( \mu\text{l} \) & frozen. One \( \mu\text{l} \) ethidium bromide was added just before use.

Denaturing gel mix for RNA:

The following were mixed in an RNase-free glass beaker and melted in a microwave;

- 800 mg agarose
- 58 ml DEPC.\( \text{H}_2\text{O} \)

The agarose was allowed to cool slightly before being placed in a fume hood and adding:
14 ml 37% formaldehyde
8 ml 10 x MOPS buffer

The reagents were mixed gently, poured into RNase-free gel casting apparatus and allowed to set in a fume hood.

**Denhardt's solution [100 x]:**
- 2% (w/v) bovine serum albumin (BSA)
- 2% (w/v) Ficoll (mol.wt. 400,000)
- 2% (w/v) polyvinylpyrrolidone (mol.wt. 400,000)

The solution was aliquoted and stored at -20° C.

**Gel loading buffer for DNA & RNA:**
- 50% glycerol
- 40 mM EDTA
- 0.2% xylene cyanol /bromophenol blue solution

The solution was aliquoted and stored at 4° C.

For use with RNA, ensure all containers are RNase-free.

**MOPS buffer [10 x]:**

For 1L of 10 x MOPS buffer, the following was prepared;
- 40 ml 2M Na Acetate (DEPC treated)
- 760 ml DEPC.H₂O
- 20.6 g MOPS powder

The pH was adjusted to 7.0 using NaOH pellets. The final volume was adjusted to 1000 ml with DEPC.H₂O and the solution was autoclaved.
PE (modified for ISH):

Use only RNase-free solutions and equipment

The following were added:

- 30.29 g Tris
- 9.31 g EDTA
- 350 ml DEPC.H$_2$O

The pH of the solution was adjusted to pH 7.5 using HCl (conc), then the following were added:

- 5 g tetra sodium pyrophosphate
- 10 g polyvinylpyrrolidone
- 10 g Ficoll

(These components may take some time to dissolve and heating to 65 °C may be required.) The volume was made up to 500 ml with DEPC.H$_2$O and the solution was stored at room temperature.

Solution D (acid GTC):

The following reagents were mixed together using clean glassware:

- 250 g Guanidinium thiocyanate (GTC)
- 293 ml sterile H$_2$O
- 17.6 ml 750 mM Na citrate (pH 7.0)
- 26.4 ml 10% SDS

This solution was then autoclaved and aliquoted x 5 ml volumes. It was stored at -20 °C for up to 12 months or at 4 °C for 6 months or at room temperature for 3 months.

To make solution D, 36 μl β-mercaptoethanol was added per 5 ml GITC solution.
A final thought...

"My experiment has not failed 1,000 times
-I have merely discovered 1,000 ways in which it does not work!"

_Thomas Edison_
REMEMBER... 1ST RULE OF DDRT-PCR

DON'T PANIC!

2ND RULE OF DDRT-PCR
IF AT FIRST YOU DON'T SUCCEED...

CHEAT!

IN CONCLUSION...

LETS DDRT-PCR MORE ICE CREAM

THE END (AT LAST!)