CLONAL EVOLUTION AND LUNG CANCER PHARMACOGENOMICS

by

TIMON PAUL HERMUS BUYS

B.Sc., The University of British Columbia, 2001

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES
(Pathology and Laboratory Medicine)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

June 2009

Abstract

Background: Lung cancer is the world's leading cause of cancer death, with a 5 year survival rate of ~16%. Several factors contribute to this poor prognosis: the limited detection of disease at treatable stages, the high metastatic potential of any primary tumour, and the variable effectiveness of chemotherapy. We applied high resolution whole genome profiling technologies to uncover genes associated with specific lung cancer phenotypes and to delineate clonal relationships between tumours.

Hypotheses: (i) Shared genetic features in tumours from the same patient are evidence of a common progenitor. (ii) Continuing clonal evolution facilitates selection for resistance genes during drug exposure. (iii) Drug response can be predicted for pre-treatment lung cancer by evaluating specific gene changes.

Materials/Methods: DNA alteration data from non-small cell lung cancers (NSCLC) were integrated with mRNA/protein expression data to identify genes contributing to tumourigenesis. Fine-mapped DNA alteration boundaries were used to evaluate clonality, discriminating multiple primary tumours from intrapulmonary metastasis. Subsequently, this approach was applied to define chemoresistance gene candidates in cells grown under drug selection. Genome alteration data for early stage lung tumours were also analyzed to define gene changes driving post-treatment recurrence in patients.

Results: We optimized collection of and genomic analysis for clinical lung cancer, identifying novel oncogene candidates (including genes contributing to tumour invasion). In addition, we successfully used DNA alteration boundaries to discriminate clonally-related tumours and define ongoing clonal evolution in both tumours and cancer cell lines, providing evidence in support of our first two hypotheses. We also identified dysregulated genes and gene pathways associated with post-treatment recurrence for clinical lung cancer. These last data suggest that chemoresistance may be an intrinsic process for the majority of cells in a pre-treatment tumour, lending support to our third hypothesis. Significantly, we also detected distinct recurrence-associated gene

changes within tumour histology subgroups, indicating that NSCLC may not be treatable as a single disease entity.

Conclusions: Global analysis of DNA alterations is an effective means for defining clonal relationships between tumours. Further, tumour phenotypes such as chemoresistance are governed by complex activation of a variety of gene networks.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Definition
AC	Adenocarcinoma
aCGH	array Comparative Genomic Hybridization
AFP	Amplified Fragment Pool
BAC	Bacterial Artificial Chromosome
bp	base pairs
BMP	Bone morphogenetic protein
CDDP	*cis*-diammine-dichloro-platinum (II), *cis*-platinum
cDNA	complementary DNA
CGH	Comparative Genomic Hybridization
CIS	Carcinoma *in situ*
CNV	Copy Number Variation
DNA	Deoxyribonucleic acid
dATP	Deoxyadenosine Triphosphate
dCTP	Deoxycytidine Triphosphate
dGTP	Deoxyguanosine Triphosphate
dNTP	Deoxynucleotide Triphosphate
dTTP	Deoxythymidine Triphosphate
DOP-PCR	Degenerate Oligonucleotide Primer PCR
ECM	Extracellular Matrix
EGFR	Epidermal Growth Factor Receptor
ESP	End sequence profiling
FFPE	Formalin-fixed, Paraffin Embedded

FISH	Fluorescence *in situ* Hybridization
LCM	Laser Capture Microdissection
LCNEC	Large Cell Neuroendocrine Carcinoma
LM-PCR	Linker-mediated PCR
LOH	Loss of Heterozygosity
Mbp	Mega-base pairs
mM	millimolar
mRNA	messenger RNA
mTOR	Mammalian Target of Rapamycin
MTT	4,5-dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide
NSCLC	Non-small Cell Lung Cancer
PCR	Polymerase Chain Reaction
RNA	Ribonucleic acid
ROMA	Representative oligonucleotide microarray analysis
SAGE	Serial Analysis of Gene Expression
SCLC	Small Cell Lung Cancer
SIGMA	System for Integrative Genomic Microarray Analysis
SNP	Single Nucleotide Polymorphism
SNR	Signal-to-noise Ratio
SDS	Sodium Dodecyl Sulfate
SqCC	Squamous Cell Carcinoma
SSC	Saline Sodium Citrate
tRNA	Transfer RNA

Acknowledgements

To complete this work, I have had to rely on the support of many generous people. I would like thank members of my thesis supervisory committee – Drs. Stephen Lam, Calum MacAulay, Victor Ling, and Colin Fyfe – who have been graciously available over the years. I would also like to thank my collaborators at the UT Southwestern Medical Center and the Ontario Cancer Institute, particularly Dr. Ming-Sound Tsao from the latter organization. Financial support from the Michael Smith Foundation for Health Research and the Canadian Institutes for Health Research has also made a significant contribution to this work.

I also owe sincere thanks to many members of the Lam lab – too many people to list here – who have been tremendous collaborators. Many of these individuals have been my co-authors or have been otherwise acknowledged in the original papers that comprise this thesis and I thank them all again for their hard work and dedication. I also owe special thanks to Drs. Lynn Mar and Steven Quayle, who have provided great counsel to me over the years.

I am of course deeply indebted to family and friends who have looked out for me during my time in graduate school. I feel particularly lucky to have parents and a brother who seem to always know when encouraging words are most needed.

In this endeavour, I owe ultimate thanks to two individuals above all others. The first is my wife Dr. Cathie Garnis, who lit the path to finishing this work and cheered my efforts along the way. The second is my supervisor, Dr. Wan Lam, who sold me on science at a point when I was ready to leave it behind and then created countless opportunities for me.

Thank you all.

Dedication

For Cath.

Co-authorship statement

Chapters 2 through 7 were originally co-authored as research manuscripts for publication. The entries below represent the complete citations for each of these works.

Chapter 2: Buys TPH, Wilson IM, Coe BP, Lockwood WW, Davies JJ, Chari R, DeLeeuw RJ, Shadeo A, MacAulay C, Lam WL. (2007) "Key Features of BAC Array Production and Usage" in *DNA Microarrays (Methods Express Series)* (Schena M, ed.), Scion Publishing, Ltd. Bloxham, Oxfordshire, UK, pp.115-145 (ISBN: 9781904842156).

Contributions: I developed the scope of this chapter and led the preparation of the manuscript and figures. Protocols and troubleshooting guides were previously defined based on the expertise of all authors and other members of the Lam lab. Section 2.8 and its associated figures were written separately for inclusion with this thesis.

Chapter 3: Garnis C, Davies JJ, Buys TP, Tsao MS, MacAulay C, Lam S, Lam WL. (2005) Chromosome 5p aberrations are early events in lung cancer: implication of glial cell line-derived neurotrophic factor in disease progression. *Oncogene*, 24(30):4806-12.

Contributions: My contribution to this work included microdissection of tumour cells from clinical lung cancer tissue, execution of mRNA expression experiments, and assistance with manuscript assembly. This work and Chapter 4 have been included in this thesis as 'proof of principle' for experimental techniques employed in later thesis chapters.

Chapter 4: Campbell JM, Lockwood WW, Buys TPH, Chari R, Coe BP, Lam S, and Lam WL. (2008) Integrative genomic and gene expression analysis of chromosome 7 identified novel oncogene loci in non-small cell lung cancer, *Genome*, 51(12):1032-9.

Contributions: I was responsible for the collection of, microdissection of, and nucleic acid isolation from the clinical lung cancer tissues used in this work. Along with JMC and WWL, I also contributed significantly to the experimental design of this work and the subsequent manuscript assembly.

Chapter 5: Buys TPH, Aviel-Ronen S, Waddell TK, Lam WL, and Tsao MS. (2008) Defining genomic alteration boundaries for a combined small cell and non-small cell lung carcinoma, *Journal of Thoracic Oncology*, 4(2): 227-39.

Contributions: I developed the question underlying this report, performed data analyses, proposed the clonal relationship between lesions, and wrote the resulting manuscript. SAR, TKW, and MST provided the case for this analysis and performed immunohistochemistry experiments.

Chapter 6: Buys TPH, Chari R, Lee EHL, Zhang M, MacAulay C, Lam S, Lam WL, Ling V. (2007) Genetic changes in the evolution of multidrug resistance for cultured human ovarian cancer cells. *Genes, Chromosomes & Cancer*, 46:1069-1079.

Contribution: I conceived the experiments, analyzed data, and wrote the manuscript for this work. MZ assisted with cell culture experiments. RC and EHLL contributed to gene expression data analysis.

Chapter 7: Buys TPH, Chae JM, Zhu CQ, Lockwood WW, Coe BP, Chari R, Girard L, Yee J, English J, Atkar-Khattra S, Hwang D, Zhang M, Dong C, Saprunoff H, Kennett JY, Ng R, Gazdar AF, Minna JD, Tsao MS, Ling V, MacAulay C, Lam S, Lam L. (2009) Genomic signatures associated with post-treatment recurrence in early stage lung cancer, *in preparation*.

Contribution: I am the primary author of the manuscript that forms this chapter, which is one of several papers being generated as part of a large-scale lung cancer pharmacogenomics project. This work involves materials and genomic data from this larger effort. JMC, CQZ, WWL, BPC, RC, LG, JY, JE, SAK, DH, MZ, CD, HS, JYK, and RN were associated with technical support aspects of this project including pathology review, collection of clinical data, sample acquisition, molecular experimentation, and data analysis. AFG, JDM, MST, VL, CM, SL, WLL are the principal investigators for the larger project and oversaw the larger project this work draws from.

1. INTRODUCTION

1.1. Lung cancer

1.1.1. Incidence and causes

Lung cancer is the leading cause of cancer death around the world. In the past decade, North America alone has seen approximately a quarter million new lung cancer cases each year and close to 200,000 annual lung cancer deaths (NCIC 2007; Jemal et al. 2008; Ruiz-Godoy et al. 2007). The major etiological factor for this disease is smoking; approximately three quarters of patients are or were tobacco smokers, with thousands of additional patients thought to develop lung tumours each year because of exposure to tobacco smoke (Repace & Lowrey 1990; Sun et al. 2007). The global incidence of certain lung cancer subtypes (which are discussed below) has changed as the nicotine content of cigarettes has changed, with compensatory smoking habits altering the location and the type of lung tumours that occur (Sun et al. 2007). Amongst non-smokers who develop lung cancer, there is a striking disparity in incidence based on sex; while more than half of female lung cancer patients are non-smokers, the same can only be said for 15% of male patients. Possible risk factors for lung cancer besides tobacco-use include exposure to other possible carcinogens (such as radon or arsenic), heritable factors (e.g. *CYP1A1* polymorphisms), and viral exposure (Sun et al. 2007).

1.1.2. Lung tumour subtypes

Histological criteria have been used to define two major subtypes of lung cancer: small-cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC) (Travis 2002). The vast majority of diagnosed cases (~80%) are of the latter subtype. These NSCLC tumours can be further parsed into broad classes that include adenocarcinomas (30% of cases), squamous cell carcinomas (30% of cases), and large cell carcinomas (~10% of cases). Additional subdivisions exist within each of these groups and other lung cancer cases may present with unspecified histological features.

While adenocarcinomas are usually detected in the peripheral airways or lung parenchyma, squamous cell carcinomas are typically found in the central airways (Travis 2002). X-ray and computed tomography (CT) scans are commonly used to detect tumours in the chest area. Determination of lung tumour subtypes is based on pathology review of cytological and histological differences. Tumours located in central airways can be diagnosed by white light bronchoscopy and biopsy (Palcic et al. 1991). For tumours found in peripheral lung regions, diagnosis may be made by bronchial washing or brushing, transthoracic needle biopsy, or resectional biopsy (Yung 2003). In all cases, biopsied tissue undergoes fixation in formalin and is then embedded in paraffin. Sections are then cut and stained for histology review by a pathologist. The remaining formalin-fixed, paraffin embedded (FFPE) tissue may be archived for future study. Specifically, these archived tumours may be suitable for molecular analyses of DNA and protein (using techniques such as fluorescence *in situ* hybridization [FISH] or immunohisotchemistry respectively).

1.1.3. Staging and prognosis for NSCLC

The overall 5-year survival for NSCLC is ~16% (NCIC 2007; Jemal et al. 2008). Differences in staging have a significant impact on prognosis.

Localized NSCLC cases comprise only ~18% of all diagnosed lung cancers and are typically treated with surgery alone or surgery plus adjuvant chemotherapy (Ries et al. 2008). Approximately half of patients diagnosed at these early stages stage live more than 5 years (Ries et al. 2008). One possible explanation for this relatively poor prognosis in early stage disease is the presence of undetectable micrometastases. Data show that adjuvant chemotherapy confers a survival advantage for this group (Arriagada et al. 2004; Douillard et al. 2006; Winton et al. 2005). For example, post-operative treatment with a *cis*-platinum and vinorelbine doublet was shown to increase 5-year survival rates by 15% compared to surgery alone (Winton et al. 2005).

When patients with locally or regionally advanced disease can tolerate it (Stage III, unresectable), both chemotherapy and radiation are given. Chemotherapy is typically a doublet regimen, as above. Patients diagnosed at this stage represent ~25% of all

cases (NCIC 2007; Jemal et al. 2008; Ries et al. 2008). Fewer than one in six people diagnosed at this stage will live past 5 years.

Advanced lung cancer (Stage IIIB, Stage IV) is typically treated with radiation and/ or chemotherapy, though the intent is palliative rather than curative. Prognosis for patients diagnosed at this stage is very poor, with median survival rates in the range of 6-10 months and five-year survival a dismal 2%. Common sites for metastases include the brain, liver, adrenal glands, bone, kidneys, and abdominal lymph nodes (Quint & Francis 1999). Although an overall survival time increase has been observed for patients treated with chemotherapy, this benefit is typically measured only in weeks or months and over half of patients fail to respond to treatment in the first place (NSCLC Collaborative Group 1995; Shepherd 1999).

In general, the poor overall prognosis seen for lung cancer patients is attributed to multiple factors including:

- a lack of sensitive methods for early diagnosis,
- the possible development of metastases when primary tumours are still small in size, and
- the absence of systemic therapies capable of dealing effectively with micrometastatic disease.

Patients with early stage, localized lung cancer show the best response to therapies and exhibit the greatest survival rate compared to patients with later stages of disease. However, based on current screening approaches, ~65% of patients present at these later stages at the time of diagnosis. In addition to arguing strongly for a greater emphasis on early detection and diagnosis, this also suggests the need to develop more effective tools for treating those patients with advanced lung tumours.

1.1.4. Genetic alterations causing disease

DNA sequence mutations, aberrant DNA methylation, changes in gene dosage, mis-expression of non-coding RNAs (e.g. microRNAs), and histone modifications can all

cause dysregulation of genes that contribute to cancer phenotypes. Several such molecular alterations have been characterized for lung cancer, including changes that can be specifically associated with disease subtypes or clinical behaviour (e.g. drug response) (Sato et al. 2007). In a clinical setting, diagnosis and treatment of lung cancer patients is largely influenced by tumour stage, tumour subtype (as defined by cell morphology), and the associated demographic data (such as the patient's age and smoking history). Molecular biomarkers may be used to gain additional information for a given case. Fluorescence *in situ* hybridization (FISH) can be used to confirm a lung cancer diagnosis based on the presence of characteristic changes in DNA copy number and immunohistochemistry (IHC) can be used to help distinguish between tumour subtypes (e.g. differences in p63 or CK5/6 staining) (Halling & Kipp 2007; Kargi et al. 2007; Strickland-Marmol et al. 2007; Varadhachary et al. 2004). Generally, these molecular approaches are used as an adjunct to conclusions based on the conventional criteria described above. Patient management strategies may prove more effective where they are guided by biomarkers that reflect the underlying gene changes driving a given tumour.

Although IHC is commonly used in the clinic and several recurrent lung cancer oncogenes and tumour suppressor gene candidates have been identified, prognostic protein biomarkers are not commonly applied as a standard approach for guiding treatment decisions. This may be due to the fact that any protein candidate that has been evaluated in at least eight different studies has produced conflicting results (though the data do seem to suggest that p53 staining can be used to predict poor prognosis in a vast majority of cases) (Zhu et al. JCP 2006). IHC remains an attractive tool given its effective adoption in the clinic, though discovery of new protein biomarkers for lung cancer may be challenging due to the limited availability of robust platforms for global proteomic analysis.

Multiple gene expression signatures have been associated with clinical parameters such as tumour histology and patient survival (Beer et al. 2002; Bhattacharjee et al. 2001; Garber et al. 2001; Parmigiani et al. 2004; Tomida et al. 2004). Unfortunately,

there is limited overlap between results from these approaches, possibly due to small contributing sample sizes, variations in sample collection techniques, differences in the microarray platform used, and myriad approaches to data analysis. Meta-analyses have derived predictive models for disease recurrence that outperform clinically-reproducible signatures from these data, though they may be less powerful when applied by other hands (Larsen et al. 2007; Potti et al. 2006). Examinations within prognostic signatures have led to refined signatures comprised of only a few genes that can be validated in independent data sets (Chen et al. 2007; Lau et al. 2007).

Several recurrent regions of genomic gain and loss have been identified in lung cancer (Sato et al. 2007). These alterations may be heritable or they may be somatic. Heritable alterations can be associated with enhanced susceptibility to lung cancer in non-smokers (Hung et al. 2003; Raimondi et al. 2005). DNA repair enzyme polymorphisms (e.g. *XRCC3*) can be associated with patient response to DNA damaging chemotherapeutic agents (Rosell et al. 2006). Somatic gene amplification and deletion represents a significant factor in lung tumourigenesis. Some alterations, such as gains on chromosome arm 5p and losses on chromosome 3p, have been documented as occurring in approximately half of NSCLC cases (Sato et al. 2007; Choi et al. 2006; Shibata et al. 2005; Tonon et al. 2005; Zhao et al. 2004; Zhu et al. 2006) (Table 1.1). Recent studies have demonstrated that specific regions of genomic alterations – and even some specific oncogenes – are associated with different NSCLC subtypes, a significant result given uniform approaches to treatment for lung cancer (Kwei et al. 2008; Petersen et al. 1997; Weir et al. 2007). Genomic alterations in *EGFR* are associated with outcome in lung cancer; data suggest mutation or amplification of this gene can be used to predict response to tyrosine kinase inhibitors such as erlotinib and gefitinib (Karamouzis et al. 2007).

In practice, integration of genomic data with additional levels of molecular analysis is an effective means for confirming the consequences of observed DNA changes. For example, genes within an amplicon that are not overexpressed are less likely to represent key causal changes than those that are exhibiting increased transcript levels

(Coe & Lockwood et al. 2006; Lockwood et al. 2008). Referring again to the example where different types of *EGFR* activation may predict sensitivity to tyrosine kinase inhibitors (TKIs), it is clear that future diagnostics and predictive biomarkers will likely be based on combined analysis from different molecular levels. It is also possible that the same molecular marker may be used to guide treatment decisions in different contexts. For example, data suggest that activation of the nucleotide excision DNA repair pathway – particularly its *ERCC1* component – is associated with favourable outcome where lung cancer patients are treated with surgery alone, but a poorer outcome where chemotherapy is also applied (Gazdar 2007). DNA is more attractive for clinical application than RNA given its relative stability in fresh and archived tissues and the existence of clinical tools based on its analysis (e.g. FISH).

1.2. Chemotherapy, chemoresistance, and clonal expansion in cancer

1.2.1. Types of chemotherapy

Tumour cells are understood to grow and divide more rapidly than non-tumour cells. Consequently, therapies targeting mitosis and DNA synthesis have been successful. For example, methotrexate, a folate metabolism antagonist, greatly improved survival in childhood leukemias when it was first used (Farber & Diamond 1948). This result is intuitive since 1) folic acid is processed in the cell to provide single carbon groups for synthesis of RNA and DNA precursors and 2) elevated rates of nucleotide synthesis would be essential in cells with higher rates of division (e.g. tumour cells). Similarly, compounds targeting microtubules (e.g. vinca alkaloids, taxanes) would be likely to disrupt cells with high rates of mitosis. On the other hand, platinum-containing compounds such as cis-diammine-dichloro-platinum (II) (CDDP) are understood to work by triggering apoptosis via irreversible DNA damage (Perez 1998).

Application of chemotherapy has been refined to improve efficacy. Doublet regimens, designed to target hyperproliferative tumour cells at multiple susceptible points, are the cornerstone of many treatment strategies. Standard approaches for chemotherapy to all stages in lung cancer often involve application of a platinum-based agent (to damage DNA) in combination with a second drug targeting a different cell process (e.g.

microtubule formation [taxol] or DNA base synthesis [gemcitabine]) (Schiller et al. 2002). Adjuvant chemotherapy plus surgery has been demonstrated to improve survival relative to surgery alone for early stage lung cancer patients (Winton et al. 2005). Palliative treatment of late stage lung cancer by chemoradiation has been demonstrated to reduce tumour burden and contribute to slightly longer overall survival (Sirzen et al. 2003; BCCA 2008).

The discovery of specific oncogenes and tumour suppressors spawned efforts to identify therapies targeted against these gene changes that were understood to drive cancer phenotypes. Imatinib, which targets the BCR-ABL fusion protein and has had a striking impact on survival rates for CML, embodies the hoped-for success in targeted thereapies (Sherbenou & Druker 2007). In lung cancer, gefitinib and erlotinib – both tyrosine kinase inhibitors – have also had demonstrable success, albeit for a specific subset of patients (Karamouzis et al. 2007).

Given that cancer is understood to manifest through the accumulation of multiple genetic alterations affecting several cell functions (Hanahan & Weinberg 2000), there is now a search for drugs with pleiotropic effects. One example of this are therapies directed against HSP90, a protein previously demonstrated to regulate factors controlling a variety of essential tumourigenic processes (including autologous growth factor signalling, anti-apoptotic activation, development of limitless replicative potential, and invasion of neighbouring tissues (Workman 2004)). In lung cancer, HSP90 inhibitors have been demonstrated to have efficacy for both *in vitro* and early clinical trials (Shimamura & Shapiro 2008).

1.2.2. Chemoresistance

Chemoresistance can limit the effective treatment of a variety of cancers. Direct action against drugs or compensation for drug effects can both drive resistance mechanisms (Harrison 1995). Examples of direct action against drugs include increasing detoxification processes, preventing conversion of drug precursors into their active states, and preventing accumulation of drugs within the cell. Activation of DNA repair and anti-apoptotic processes, alterations to the amounts and activities of drug targets,

and activation of analogous pathways not targeted by drugs can all serve as mechanisms for compensating for drug effects. It is possible that genetic alterations driving chemoresistance may be induced by drug, though experimental data have supported a model where these alterations arose in a subpopulation of cells prior to treatment and that these cells then expanded following drug selection (Lederberg & Lederberg 1952; Luria & Delbruck 1943; Goldie 2001). For many types of chemotherapy, including those commonly applied to lung cancer, multiple resistance mechanisms may exist. DNA-damaging CDDP may be countered either by activation of DNA repair pathway members (e.g. *ERCC1* or *RRM1*) or by activation of anti-apoptotic signals (e.g. *BCL2*) that limit the usual apoptotic consequences of DNA adduct formation (Perez 1998). Anti-mitotic vinca alkaloids may be countered by mutations in tubulin targets, changes in the ratio of tubulin isotypes, or by activation of transmembrane ABC drug transporter pumps which exclude drug from cells before they can have an effect (Dumontet & Sikic 1999).

1.3. Clonal evolution

1.3.1. Clonal evolution of tumours

Tumour cell populations are heterogeneous, with intratumoural genetic variation a product of ongoing evolution from a precursor diseased cell. Genomic divergence during tumourigenesis is a product of exposure to cancer-causing agents, the rate of cell division, and the activity of apoptotic and DNA repair machinery. Ultimately, cells that evolve alterations conferring a survival advantage will expand more rapidly and come to comprise the majority of a given tumour mass.

1.3.2. Determining whether tumours are clonally-related or represent multiple primary lesions

Patients can present with multiple tumours (“synchronous” or “metachronous” lesions). It is important to distinguish cases of multiple primary cancers from cases where there is a shared progenitor (i.e. intrapulmonary metastasis). The frequency of multiple primary tumours varies among cancer types: 3-5% for breast tumours, >30% in prostate and

~20% in hepatocellular cancer (Demandante et al. 2003; Imyanitov et al. 2002; Matsumoto et al. 2001). For lung cancer, incidence of synchronous tumours is ~2% (Martini & Melamed 1975). Establishing the relationship between such tumours is not only essential for understanding the underlying tumour biology, it will also impact disease staging and patient management.

Clinical diagnosis of multiple primary tumours typically relies on differences in location, histology, and staging. Unfortunately, these criteria may not reflect the genetic reality underlying disease behavior; histologically similar synchronous tumours may exhibit genetic evidence of distinct clonal origins or histologically distinct lesions could have a shared progenitor. Previous efforts to delineate clonality by non-histopathological means have relied on evaluation of only a few genetic alterations (examples including the mutational status of individual oncogenes or tumour suppressors [e.g. *p53* or *K-Ras*], X chromosome inactivation, and DNA methylation changes) (Chang et al. 2007; Dacic et al. 2005; Lau et al. 1997; Wain et al. 1986). Recent applications of multi-locus assays have offered a more detailed description of the similarities and differences between synchronous tumours (D'Adda et al. 2008; Fellegara et al. 2008; Froio et al. 2008; Huang et al. 2002; Murase et al. 2003). Whole genome technologies have allowed fine characterization of genome alterations, providing compelling evidence for shared clonal origins where identical complex rearrangements have been apparent in synchronous lung tumours (Gallegos Ruiz et al. 2007; Takahashi et al. 2007) (see also Chapter 5).

1.4. Technologies for cancer genome comparisons

Whole genome analysis of tumour samples is typically based on detection of regions of allelic imbalance or loss of heterozygosity, evaluation of chromosomal aberrations and re-arrangements by molecular cytogenetic techniques, or identification of segmental DNA copy number alterations to identify key genomic features contributing to disease phenotypes (Garnis et al. 2004) (Figure 1.1).

1.4.1. Loss of heterozygosity

LOH is detected by microsatellite analysis of simple sequence repeat (SSR) polymorphisms. For an individual who is heterozygous at a given allele, PCR analysis using primers flanking a specific SSR should yield two signals (one for each allele). When the signal intensity ratio for alleles seen in a tumour differs from what is seen in a matched normal specimen, LOH is inferred. Microarray platforms designed for parallel analysis of several single nucleotide polymorphisms (SNPs) allows high resolution detection of both genotype and relative gene copy number changes (Zhao et al. 2004; Lockwood & Chari et al. 2006).

1.4.2. Cytogenetics

Molecular cytogenetics techniques (e.g. G-banding and spectral karyotyping [SKY]) survey the whole genome to detect DNA ploidy changes and chromosomal rearrangements. G-banding uses staining of metaphase chromosome spreads to detect rearrangements or gain/ loss of chromosome bands. SKY involves the generation of a virtual karyogram by use of differentially labeled chromosome-specific probes; representation of each chromosome by a different colour allows detection of chromosomal rearrangements (Bayani & Squire 2002). These approaches have been widely used in clinical settings, particularly for hematological malignancies because they are less karyotypically complex than carcinomas.

Fluorescence *in situ* hybridization (FISH) is an approach that evaluates alterations at a specific genome locus. This approach offers the advantage that it can provide data on a cell-by-cell basis, overcoming tissue heterogeneity that can mask alteration features specific to a cell subpopulation. Gain or loss of hybridization signals indicate DNA duplication and deletion respectively. Split signals denote a translocation event. Multi-colour FISH (M-FISH), which is based on probes that fluoresce at different wavelengths, allows parallel examination of several genomic loci at once (Gray et al. 1991).

1.4.2.1. Comparative Genomics Hybridization

Segmental DNA copy number gains and losses may be detected by comparative genomic hybridization (CGH). In CGH, tumour and reference DNA samples are

differentially labeled, mixed, and then allowed to competitively hybridize to metaphase chromosomes. Whole genome copy number profiles are inferred from the signal ratio between the two labels (Kallioniemi et al. 1992).

Adaptation of CGH to arrayed DNA targets rather than cells arrested in metaphase has allowed for enhanced resolution of this technology (Solinas-Toldo et al. 1997; Lockwood et al. 2006). The earliest such analyses were performed on cDNA microarrays (Pollack et al. 1999). Array platforms comprised of large insert clones (e.g. bacterial artificial chromosomes [BACs]) allowed interrogation of unannotated genes, regulatory regions, and intergenic sequences (Snijders et al. 2001; Greshock et al. 2004). The increased detection sensitivity and reduced DNA input requirements of these platforms have facilitated analysis of clinical tumour tissues, which may have lower DNA yields and poorer DNA quality (both of these issues being a particular challenge when working with formalin fixed paraffin embedded [FFPE] tissues from hospital archives) (Lockwood et al. 2006). Analysis with oligonucleotide-based platforms, such as those used for single-nucleotide polymorphism (SNP) analysis and representative oligonucleotide microarray analysis (ROMA), offer marked improvements in the number of loci interrogated in a single experiment relative to earlier platforms (Lucito et al. 2003; Matsuzaki et al. 2004). As mentioned above, SNP arrays facilitate both the detection of LOH and DNA copy number changes (Bignell et al. 2004; Zhao et al. 2004) (though some SNP loci will be uninformative for allelic status due to homozygosity). Selection of an array CGH platform is best dictated by the samples being analyzed, with some of the major considerations already discussed elsewhere: the input quality and quantity of available DNA, the ability to detect genetic alterations in heterogeneous samples, and the ability to detect focal alterations (as influenced by the distribution of array elements) (Davies et al. 2005; Garnis & Coe et al. 2005; Coe et al. 2008; Coe et al. 2007).

1.4.3. DNA sequencing-based technologies

Genome sequencing is increasingly being applied to a variety of cancer types. Digital karyotyping – a technique analogous serial analysis of gene expression (SAGE), except genomic DNA is used to generate concatenated DNA tags for sequence analysis

(Velculescu et al. 1995) – deduces relative DNA copy number by enumerating sequence tags representing loci throughout the genome (Wang et al. 2002). This approach has been used to uncover activating alterations for ovarian tumours and has been adapted to assess epigenetic alterations in breast tumours (Park et al. 2006; Shih et al. 2005; Bloushtain-Qimron et al. 2008). Tumour genomes are represented in fosmid or BAC clones with End sequence profiling (ESP), a technique wherein copy number changes and chromosomal rearrangements are identified by sampling of clones by end sequencing (Volik et al. 2003; Tuzun et al. 2005; Volik et al. 2006). Sequencing-based analyses have also been undertaken for mutational analysis of tumour genomes. Recently, the mutational status for >13,000 protein encoding genes was determined for individual colorectal and breast tumours by a Sanger-sequencing-based approach (Sjoblom et al. 2006). Recurring mutations (nonsense mutations, missense mutations, etc.) were identified at hundreds of novel candidate loci, underscoring the complexity of tumourigenic processes. Similarly broad mutational studies have been undertaken for a variety of tumour types (Greenman et al. 2007; Jones et al. 2008; Ley et al. 2008). Such studies will increase in number as emerging technologies promise reduced costs and increased speed (e.g. pyrosequencing, multiplex polony sequencing) (Shendure et al. 2005; Metzker 2005; Costabile et al. 2006; Velculescu 2008).

1.4.4. Integration of multi-dimensional genomic data

Dysregulation in cancer cells occurs at many levels, meaning that genomic analysis using multiple complementary platforms will provide a more comprehensive description of the tumour genome. For example, an integrative study identifying alterations in DNA and mRNA expression patterns uncovered causal genetic events and their downstream effects (Pollack et al. 2002). Similarly, matching DNA copy number status with DNA methylation profiles may identify genes disrupted in both alleles and predict silencing of gene expression (Wilson et al. 2006). The need for multi-dimensional profiling of tumours has prompted the development of integrative software catered to the display and analysis of complementary datasets. Programs such as Magellan, ACE-it, and VAMP are able to integrate DNA alteration and gene expression data (van Wieringen et al. 2006; Kingsley et al. 2006; Rosa et al. 2006), while recently developed SIGMA

(System for Integrative Genomic Microarray Analysis) is a user interface for direct mining of multi-dimensional data (Chari & Lockwood et al. 2006). The ability to merge data from various genomic profiling platforms will facilitate cancer gene discovery and contribute to the understanding of the underlying causes for the diversity of existing cancer phenotypes.

1.5. Hypotheses and objectives

The work in this thesis was founded upon the following hypotheses:

1. The clonal relationship between tumours from the same patient is reflected in their shared genetic features.
2. Ongoing clonal evolution facilitates resistance mechanisms following drug exposure.
3. Prediction of drug response can be made based on specific genetic features in pre-treatment lung tumours.

These hypotheses result in the following study objectives:

1. To delineate clonal relationships between lung tumours from the same patient.
2. To demonstrate clonal evolution under drug selection in model systems.
3. To deduce genomic signatures predictive of drug response in lung cancer.

1.6. Specific aims and thesis outline

The overall framework for this thesis is detailed below. This work represents a collection of manuscripts that have been assembled to address the above hypotheses.

Aim 1 – Optimization of approaches to clinical sample collection for the generation of genomic profiles for lung tumours (Chapters 2-4).

Chapter 2 provides a rigorous description of the tiling path array CGH platform that was used in the genomic analyses described by all subsequent chapters. This description

includes details about array production, considerations regarding experimental design and the use of different sample types, and approaches to analyzing the resulting data. It also provides a primer on the collection procedures that were used when handling cases of clinical lung cancer, including the methods used to optimize analysis of lung tumour cells only.

When this work was started, successful application of the tiling set array CGH platform to lung cancer tissue had not previously been described. Moreover, the wealth of data generated by this platform required additional levels of analysis beyond what was described in Chapter 2. Chapters 3 and 4 describe application of the CGH platform to DNA isolated from both cell lines and clinical specimens and details additional analytical approaches that were used to define significant genomic alteration features in each data set. These two chapters also describe validation of genomic alterations of clinical significance by additional molecular analyses. Chapters 2-4 are 'proof of principle' works that helped to develop tools for addressing the hypotheses and objectives listed above.

Aim 2 – Using segmental genomic alteration boundaries as signature markers for clonality where a patient presents with synchronous lung tumours (Chapter 5).

In Chapter 5, the approaches to lung tumour genome analysis that were described earlier were applied to delineate whether multiple tumours from the same individual could have a shared progenitor. Because pulmonary metastases and multiple independent primary tumours are staged and managed differently in the clinic, it is essential to determine the clonal relationship between tumours found in the same patient. We simultaneously tracked alteration features throughout tumour genomes and aligned high resolution genomic profiles for synchronous lung tumours to identify shared genome alteration boundaries. Where these rare precisely aligned boundaries were the same, we would have a means of inferring a shared clonal origin. Data presented in this chapter directly support our first hypothesis and addressed our first objective.

Aim 3 – Characterizing drug resistance features emerging during drug-selected clonal evolution (Chapter 6).

The approaches used to define shared genomic alterations in tumours emerging from a common progenitor were also applied to define regions of difference in populations of cells undergoing clonal evolution. In Chapter 6, genomic and gene expression profiling technologies were used to identify factors driving the emergence of chemoresistance for cells grown under drug selection. Data for drug sensitive cells and a series of resistant derivatives were compared. Global genomic and gene expression technologies were used to determine whether emerging chemoresistance was a product of a single factor with steadily increasing activation or whether it was the result of multiple factors. Our second hypothesis and objective were directly addressed by the work presented in Chapter 6.

Aim 4 – Identification of genomic alterations in pretreatment lung tumours that can be associated with drug response (Chapter 7).

The work described in Chapter 7 was undertaken to define genome features in NSCLC tumours that were predictive of response to chemotherapy. Experimental and analytical approaches from earlier chapters were combined with clinical outcome data to define key alterations in this panel of lung cancer cases. This work directly addresses our third hypothesis and objective. In successfully identifying predictive features for this panel of samples, the data from this chapter also facilitates comment on the behaviour of clonal populations of lung cancer cells under drug selection.

Table 1.1 – Previously reported genome loci commonly altered in non-small cell lung cancer.

	Gain	Loss
Regions of chromosomal alteration	1q21-q25, 5p, 7q11.2, 9q34, 11q12-q13, 14q11-q14, 15q, 17q25, 20q	2q, 3p, 4p15-p16, 6q24, 8p22-p23, 9p, 10q21, 18q, 21q22, 22q

Figure 1.1

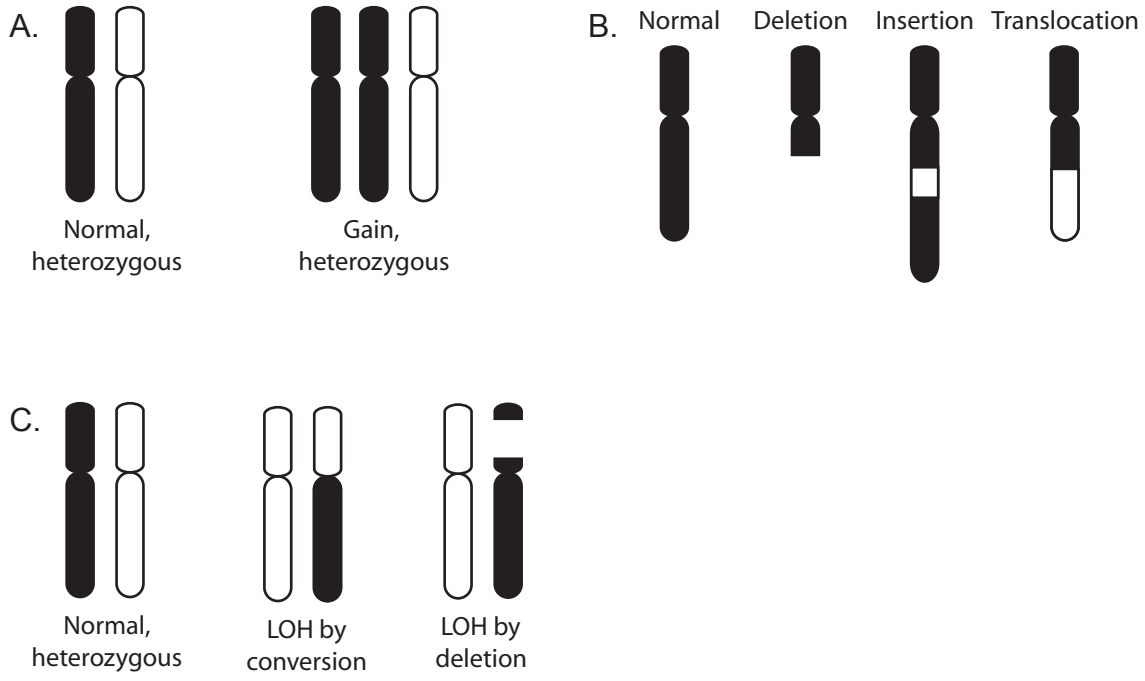


Figure 1.1 Types of genomic aberration. A. Alterations affecting normal allelic balance and DNA dosage. One of the common genomic alterations which may occur is the generation of an aneuploid or polyploid genome through gain or loss of chromosomes. This can be detected by copy number sensitive technologies such as CGH, quantitative PCR, and cytogenetic evaluation. B. Segmental Copy Number Alterations. DNA copy number alterations and structural re-arrangements are commonly observed in cancer genomes, which affect only part of a chromosome. These may include the loss of DNA material, duplication of chromosomal segments, or translocation of chromosomal ends by recombination. C. Allelic imbalance. LOH can arise from a deletion event or gene conversion during mitosis.

1.7. References

Arriagada, R., B. Bergman, A. Dunant, T. Le Chevalier, J. P. Pignon & J. Vansteenkiste, 2004. Cisplatin-based adjuvant chemotherapy in patients with completely resected non-small-cell lung cancer. *N Engl J Med,* 350(4), 351-60.

Bayani, J. M. & J. A. Squire, 2002. Applications of SKY in cancer cytogenetics. *Cancer Invest,* 20(3), 373-86.

Beer, D. G., S. L. Kardia, C. C. Huang, T. J. Giordano, A. M. Levin, D. E. Misek, L. Lin, G. Chen, T. G. Gharib, D. G. Thomas, M. L. Lizyness, R. Kuick, S. Hayasaka, J. M. Taylor, M. D. Iannettoni, M. B. Orringer & S. Hanash, 2002. Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat Med,* 8(8), 816-24.

Bhattacharjee, A., W. G. Richards, J. Staunton, C. Li, S. Monti, P. Vasa, C. Ladd, J. Beheshti, R. Bueno, M. Gillette, M. Loda, G. Weber, E. J. Mark, E. S. Lander, W. Wong, B. E. Johnson, T. R. Golub, D. J. Sugarbaker & M. Meyerson, 2001. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci U S A,* 98(24), 13790-5.

Bignell, G. R., J. Huang, J. Greshock, S. Watt, A. Butler, S. West, M. Grigorova, K. W. Jones, W. Wei, M. R. Stratton, P. A. Futreal, B. Weber, M. H. Shapero & R. Wooster, 2004. High-resolution analysis of DNA copy number using oligonucleotide microarrays. *Genome Res,* 14(2), 287-95.

Bloushtain-Qimron, N., J. Yao, E. L. Snyder, M. Shipitsin, L. L. Campbell, S. A. Mani, M. Hu, H. Chen, V. Ustyansky, J. E. Antosiewicz, P. Argani, M. K. Halushka, J. A. Thomson, P. Pharoah, A. Porgador, S. Sukumar, R. Parsons, A. L. Richardson, M. R. Stampfer, R. S. Gelman, T. Nikolskaya, Y. Nikolsky & K. Polyak, 2008. Cell type-specific DNA methylation patterns in the human breast. *Proc Natl Acad Sci U S A,* 105(37), 14076-81.

British Columbia Cancer Agency, Cancer Management Guidelines for Non-small Cell Lung Cancer, (2008). http://www.bccancer.bc.ca/HPI/CancerManagementGuidelines/Lung/6ManagementPolicies/641NonSmallCellLungCancer/13PalliativeRadiotherapyforLocallyAdvancedandMetastaticNSCLS.htm.

Chang, Y. L., C. T. Wu, S. C. Lin, C. F. Hsiao, Y. S. Jou & Y. C. Lee, 2007. Clonality and prognostic implications of p53 and epidermal growth factor receptor somatic aberrations in multiple primary lung cancers. *Clin Cancer Res,* 13(1), 52-8.

Chari, R., Lockwood, WW. Coe, BP. Chu, A. Macey, D. Thomson, A. Davies, JJ. MacAulay, C. Lam, WL., 2006. SIGMA: A System for Integrative Genomic

Microarray Analysis of Cancer Genomes. *Genome Biology,* submitted(submitted).

Chen, H. Y., S. L. Yu, C. H. Chen, G. C. Chang, C. Y. Chen, A. Yuan, C. L. Cheng, C. H. Wang, H. J. Terng, S. F. Kao, W. K. Chan, H. N. Li, C. C. Liu, S. Singh, W. J. Chen, J. J. Chen & P. C. Yang, 2007. A five-gene signature and clinical outcome in non-small-cell lung cancer. *N Engl J Med,* 356(1), 11-20.

Choi, J. S., L. T. Zheng, E. Ha, Y. J. Lim, Y. H. Kim, Y. P. Wang & Y. Lim, 2006. Comparative genomic hybridization array analysis and real-time PCR reveals genomic copy number alteration for lung adenocarcinomas. *Lung,* 184(6), 355-62.

Coe, B. P., W. W. Lockwood, L. Girard, R. Chari, C. Macaulay, S. Lam, A. F. Gazdar, J. D. Minna & W. L. Lam, 2006. Differential disruption of cell cycle pathways in small cell and non-small cell lung cancer. *Br J Cancer,* 94(12), 1927-35.

Coe, B. P., C. Macaulay, W. L. Lam, B. Ylstra, B. Carvalho & G. A. Meijer, 2008. Comment re: a comparison of DNA copy number profiling platforms. *Cancer Res,* 68(10), 4010; author reply

Coe, B. P., B. Ylstra, B. Carvalho, G. A. Meijer, C. Macaulay & W. L. Lam, 2007. Resolving the resolution of array CGH. *Genomics,* 89(5), 647-53.

Costabile, M., A. Quach & A. Ferrante, 2006. Molecular approaches in the diagnosis of primary immunodeficiency diseases. *Hum Mutat*.

D'Adda, T., G. Pelosi, C. Lagrasta, C. Azzoni, L. Bottarelli, S. Pizzi, I. Troisi, G. Rindi & C. Bordi, 2008. Genetic alterations in combined neuroendocrine neoplasms of the lung. *Mod Pathol,* 21(4), 414-22.

Dacic, S., D. N. Ionescu, S. Finkelstein & S. A. Yousem, 2005. Patterns of allelic loss of synchronous adenocarcinomas of the lung. *Am J Surg Pathol,* 29(7), 897-902.

Davies, J. J., I. M. Wilson & W. L. Lam, 2005. Array CGH technologies and their applications to cancer genomes. *Chromosome Res,* 13(3), 237-48.

Demandante, C. G., D. A. Troyer & T. P. Miles, 2003. Multiple primary malignant neoplasms: case report and a comprehensive review of the literature. *Am J Clin Oncol,* 26(1), 79-83.

Douillard, J. Y., R. Rosell, M. De Lena, F. Carpagnano, R. Ramlau, J. L. Gonzales-Larriba, T. Grodzki, J. R. Pereira, A. Le Groumellec, V. Lorusso, C. Clary, A. J. Torres, J. Dahabreh, P. J. Souquet, J. Astudillo, P. Fournel, A. Artal-Cortes, J. Jassem, L. Koubkova, P. His, M. Riggi & P. Hurteloup, 2006. Adjuvant

vinorelbine plus cisplatin versus observation in patients with completely resected stage IB-IIIA non-small-cell lung cancer (Adjuvant Navelbine International Trialist Association [ANITA]): a randomised controlled trial. *Lancet Oncol,* 7(9), 719-27.

Dumontet, C. & B. I. Sikic, 1999. Mechanisms of action of and resistance to antitubulin agents: microtubule dynamics, drug transport, and cell death. *J Clin Oncol,* 17(3), 1061-70.

Farber, S. & L. K. Diamond, 1948. Temporary remissions in acute leukemia in children produced by folic acid antagonist, 4-aminopteroyl-glutamic acid. *N Engl J Med,* 238(23), 787-93.

Fellegara, G., T. D'Adda, F. P. Pilato, E. Froio, L. Ampollini, M. Rusca & G. Rindi, 2008. Genetics of a combined lung small cell carcinoma and large cell neuroendocrine carcinoma with adenocarcinoma. *Virchows Arch,* 453(1), 107-15.

Froio, E., T. D'Adda, G. Fellegara, L. Ampollini, P. Carbognani & G. Rindi, 2008. Three different synchronous primary lung tumours: a case report with extensive genetic analysis and review of the literature. *Lung Cancer,* 59(3), 395-402.

Gallegos Ruiz, M. I., H. van Cruijsen, E. F. Smit, K. Grunberg, G. A. Meijer, J. A. Rodriguez, B. Ylstra & G. Giaccone, 2007. Genetic heterogeneity in patients with multiple neoplastic lung lesions: a report of three cases. *J Thorac Oncol,* 2(1), 12-21.

Garber, M. E., O. G. Troyanskaya, K. Schluens, S. Petersen, Z. Thaesler, M. Pacyna-Gengelbach, M. van de Rijn, G. D. Rosen, C. M. Perou, R. I. Whyte, R. B. Altman, P. O. Brown, D. Botstein & I. Petersen, 2001. Diversity of gene expression in adenocarcinoma of the lung. *Proc Natl Acad Sci U S A,* 98(24), 13784-9.

Garnis, C., T. P. Buys & W. L. Lam, 2004. Genetic alteration and gene expression modulation during cancer progression. *Mol Cancer,* 3, 9.

Garnis, C., B. P. Coe, S. L. Lam, C. Macaulay & W. L. Lam, 2005. High-resolution array CGH increases heterogeneity tolerance in the analysis of clinical samples. *Genomics,* 85(6), 790-3.

Gazdar, A. F., 2007. DNA repair and survival in lung cancer--the two faces of Janus. *N Engl J Med,* 356(8), 771-3.

Goldie, J. H., 2001. Drug resistance in cancer: a perspective. *Cancer Metastasis Rev,* 20(1-2), 63-8.

Gray, J. W., J. Lucas, O. Kallioniemi, A. Kallioniemi, W. L. Kuo, T. Straume, D. Tkachuk, T. Tenjin, H. U. Weier & D. Pinkel, 1991. Applications of fluorescence in situ hybridization in biological dosimetry and detection of disease-specific chromosome aberrations. *Prog Clin Biol Res,* 372, 399-411.

Greenman, C., P. Stephens, R. Smith, G. L. Dalgliesh, C. Hunter, G. Bignell, H. Davies, J. Teague, A. Butler, C. Stevens, S. Edkins, S. O'Meara, I. Vastrik, E. E. Schmidt, T. Avis, S. Barthorpe, G. Bhamra, G. Buck, B. Choudhury, J. Clements, J. Cole, E. Dicks, S. Forbes, K. Gray, K. Halliday, R. Harrison, K. Hills, J. Hinton, A. Jenkinson, D. Jones, A. Menzies, T. Mironenko, J. Perry, K. Raine, D. Richardson, R. Shepherd, A. Small, C. Tofts, J. Varian, T. Webb, S. West, S. Widaa, A. Yates, D. P. Cahill, D. N. Louis, P. Goldstraw, A. G. Nicholson, F. Brasseur, L. Looijenga, B. L. Weber, Y. E. Chiew, A. DeFazio, M. F. Greaves, A. R. Green, P. Campbell, E. Birney, D. F. Easton, G. Chenevix-Trench, M. H. Tan, S. K. Khoo, B. T. Teh, S. T. Yuen, S. Y. Leung, R. Wooster, P. A. Futreal & M. R. Stratton, 2007. Patterns of somatic mutation in human cancer genomes. *Nature,* 446(7132), 153-8.

Greshock, J., T. L. Naylor & A. Margolin, 2004. 1-Mb resolution array-based comparative genomic hybridization using a BAC clone set optimized for cancer gene analysis. *Genome Res,* 14(1), 179-87.

Halling, K. C. & B. R. Kipp, 2007. Fluorescence in situ hybridization in diagnostic cytology. *Hum Pathol,* 38(8), 1137-44.

Hanahan, D. & R. A. Weinberg, 2000. The hallmarks of cancer. *Cell,* 100(1), 57-70.

Harrison, D. J., 1995. Molecular mechanisms of drug resistance in tumours. *J Pathol,* 175(1), 7-12.

Huang, J., C. Behrens, Wistuba, II, A. F. Gazdar & J. Jagirdar, 2002. Clonality of combined tumors. *Arch Pathol Lab Med,* 126(4), 437-41.

Hung, R. J., P. Boffetta, J. Brockmoller, D. Butkiewicz, I. Cascorbi, M. L. Clapper, S. Garte, A. Haugen, A. Hirvonen, S. Anttila, I. Kalina, L. Le Marchand, S. J. London, A. Rannug, M. Romkes, J. Salagovic, B. Schoket, L. Gaspari & E. Taioli, 2003. CYP1A1 and GSTM1 genetic polymorphisms and lung cancer risk in Caucasian non-smokers: a pooled analysis. *Carcinogenesis,* 24(5), 875-82.

Imyanitov, E. N., E. N. Suspitsin, M. Y. Grigoriev, A. V. Togo, E. Kuligina, E. V. Belogubova, K. M. Pozharisski, E. A. Turkevich, C. Rodriquez, C. J. Cornelisse, K. P. Hanson & C. Theillet, 2002. Concordance of allelic imbalance profiles in synchronous and metachronous bilateral breast carcinomas. *Int J Cancer,* 100(5), 557-64.

Jemal, A., M. J. Thun, L. A. Ries, H. L. Howe, H. K. Weir, M. M. Center, E. Ward, X. C. Wu, C. Eheman, R. Anderson, U. A. Ajani, B. Kohler & B. K. Edwards, 2008. Annual Report to the Nation on the Status of Cancer, 1975-2005, Featuring Trends in Lung Cancer, Tobacco Use, and Tobacco Control. *J Natl Cancer Inst*.

Jones, S., X. Zhang, D. W. Parsons, J. C. Lin, R. J. Leary, P. Angenendt, P. Mankoo, H. Carter, H. Kamiyama, A. Jimeno, S. M. Hong, B. Fu, M. T. Lin, E. S. Calhoun, M. Kamiyama, K. Walter, T. Nikolskaya, Y. Nikolsky, J. Hartigan, D. R. Smith, M. Hidalgo, S. D. Leach, A. P. Klein, E. M. Jaffee, M. Goggins, A. Maitra, C. Iacobuzio-Donahue, J. R. Eshleman, S. E. Kern, R. H. Hruban, R. Karchin, N. Papadopoulos, G. Parmigiani, B. Vogelstein, V. E. Velculescu & K. W. Kinzler, 2008. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science,* 321(5897), 1801-6.

Kallioniemi, A., O. P. Kallioniemi, D. Sudar, D. Rutovitz, J. W. Gray, F. Waldman & D. Pinkel, 1992. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science,* 258(5083), 818-21.

Karamouzis, M. V., J. R. Grandis & A. Argiris, 2007. Therapies directed against epidermal growth factor receptor in aerodigestive carcinomas. *Jama,* 298(1), 70-82.

Kargi, A., D. Gurel & B. Tuna, 2007. The diagnostic value of TTF-1, CK 5/6, and p63 immunostaining in classification of lung carcinomas. *Appl Immunohistochem Mol Morphol,* 15(4), 415-20.

Kingsley CB, K. W., Polikoff D, Berchuck A, Gray JW, Jain AN, 2006. Magellan: A Web Based System for the Integrated Analysis of Heterogeneous Biological Data and Annotations; Application to DNA Copy Number and Expression Data in Ovarian Cancer. *Cancer Informatics,* 2, 10-21.

Kwei, K. A., Y. H. Kim, L. Girard, J. Kao, M. Pacyna-Gengelbach, K. Salari, J. Lee, Y. L. Choi, M. Sato, P. Wang, T. Hernandez-Boussard, A. F. Gazdar, I. Petersen, J. D. Minna & J. R. Pollack, 2008. Genomic profiling identifies TITF1 as a lineage-specific oncogene amplified in lung cancer. *Oncogene,* 27(25), 3635-40.

Larsen, J. E., K. M. Fong & N. K. Hayward, 2007. Refining prognosis in non-small-cell lung cancer. *N Engl J Med,* 356(2), 190; author reply -1.

Lau, D. H., B. Yang, R. Hu & J. R. Benfield, 1997. Clonal origin of multiple lung cancers: K-ras and p53 mutations determined by nonradioisotopic single-strand conformation polymorphism analysis. *Diagn Mol Pathol,* 6(4), 179-84.

Lau, S. K., P. C. Boutros, M. Pintilie, F. H. Blackhall, C. Q. Zhu, D. Strumpf, M. R. Johnston, G. Darling, S. Keshavjee, T. K. Waddell, N. Liu, D. Lau, L. Z. Penn, F.

A. Shepherd, I. Jurisica, S. D. Der & M. S. Tsao, 2007. Three-gene prognostic classifier for early-stage non small-cell lung cancer. *J Clin Oncol,* 25(35), 5562-9.

Lederberg, J. & E. M. Lederberg, 1952. Replica plating and indirect selection of bacterial mutants. *J Bacteriol,* 63(3), 399-406.

Ley, T. J., E. R. Mardis, L. Ding, B. Fulton, M. D. McLellan, K. Chen, D. Dooling, B. H. Dunford-Shore, S. McGrath, M. Hickenbotham, L. Cook, R. Abbott, D. E. Larson, D. C. Koboldt, C. Pohl, S. Smith, A. Hawkins, S. Abbott, D. Locke, L. W. Hillier, T. Miner, L. Fulton, V. Magrini, T. Wylie, J. Glasscock, J. Conyers, N. Sander, X. Shi, J. R. Osborne, P. Minx, D. Gordon, A. Chinwalla, Y. Zhao, R. E. Ries, J. E. Payton, P. Westervelt, M. H. Tomasson, M. Watson, J. Baty, J. Ivanovich, S. Heath, W. D. Shannon, R. Nagarajan, M. J. Walter, D. C. Link, T. A. Graubert, J. F. DiPersio & R. K. Wilson, 2008. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature,* 456(7218), 66-72.

Lockwood, W. W., R. Chari, B. Chi & W. L. Lam, 2006. Recent advances in array comparative genomic hybridization technologies and their applications in human genetics. *Eur J Hum Genet,* 14(2), 139-48.

Lockwood, W. W., R. Chari, B. P. Coe, L. Girard, C. Macaulay, S. Lam, A. F. Gazdar, J. D. Minna & W. L. Lam, 2008. DNA amplification is a ubiquitous mechanism of oncogene activation in lung and other cancers. *Oncogene,* 27(33), 4615-24.

Lucito, R., J. Healy, J. Alexander, A. Reiner, D. Esposito, M. Chi, L. Rodgers, A. Brady, J. Sebat, J. Troge, J. A. West, S. Rostan, K. C. Nguyen, S. Powers, K. Q. Ye, A. Olshen, E. Venkatraman, L. Norton & M. Wigler, 2003. Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation. *Genome Res,* 13(10), 2291-305.

Luria, S. E. & M. Delbruck, 1943. Mutations of Bacteria from Virus Sensitivity to Virus Resistance. *Genetics,* 28(6), 491-511.

Martini, N. & M. R. Melamed, 1975. Multiple primary lung cancers. *J Thorac Cardiovasc Surg,* 70(4), 606-12.

Matsumoto, Y., H. Fujii, M. Matsuda & H. Kono, 2001. Multicentric occurrence of hepatocellular carcinoma: diagnosis and clinical significance. *J Hepatobiliary Pancreat Surg,* 8(5), 435-40.

Matsuzaki, H., S. Dong, H. Loi, X. Di, G. Liu, E. Hubbell, J. Law, T. Berntsen, M. Chadha, H. Hui, G. Yang, G. C. Kennedy, T. A. Webster, S. Cawley, P. S. Walsh, K. W. Jones, S. P. Fodor & R. Mei, 2004. Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. *Nat Methods,* 1(2), 109-11.

Metzker, M. L., 2005. Emerging technologies in DNA sequencing. *Genome Res,* 15(12), 1767-76.

Murase, T., H. Takino, S. Shimizu, H. Inagaki, H. Tateyama, E. Takahashi, H. Matsuda & T. Eimoto, 2003. Clonality analysis of different histological components in combined small cell and non-small cell carcinoma of the lung. *Hum Pathol,* 34(11), 1178-84.

National Cancer Institute of Canada, 2007. *Canadian Cancer Statistics 2007,* Toronto, Canada.

Non-small Cell Lung Cancer Collaborative Group. Chemotherapy in non-small cell lung cancer: a meta-analysis using updated data on individual patients from 52 randomised clinical trials. *BMJ,* 311(7010), 899-909.

Palcic, B., S. Lam, J. Hung & C. MacAulay, 1991. Detection and localization of early lung cancer by imaging techniques. *Chest,* 99(3), 742-3.

Park, J. T., M. Li, K. Nakayama, T. L. Mao, B. Davidson, Z. Zhang, R. J. Kurman, C. G. Eberhart, M. Shih Ie & T. L. Wang, 2006. Notch3 gene amplification in ovarian cancer. *Cancer Res,* 66(12), 6312-8.

Parmigiani, G., E. Garrett, R. Anbazhagan & E. Gabrielson, 2004. Molecular classification of lung cancer: a cross-platform comparison of gene expression data sets. *Chest,* 125(5 Suppl), 103S.

Perez, R. P., 1998. Cellular and molecular determinants of cisplatin resistance. *Eur J Cancer,* 34(10), 1535-42.

Petersen, I., M. Bujard, S. Petersen, G. Wolf, A. Goeze, A. Schwendel, H. Langreck, K. Gellert, M. Reichel, K. Just, S. du Manoir, T. Cremer, M. Dietel & T. Ried, 1997. Patterns of chromosomal imbalances in adenocarcinoma and squamous cell carcinoma of the lung. *Cancer Res,* 57(12), 2331-5.

Pollack, J. R., C. M. Perou, A. A. Alizadeh, M. B. Eisen, A. Pergamenschikov, C. F. Williams, S. S. Jeffrey, D. Botstein & P. O. Brown, 1999. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat Genet,* 23(1), 41-6.

Pollack, J. R., T. Sorlie, C. M. Perou, C. A. Rees, S. S. Jeffrey, P. E. Lonning, R. Tibshirani, D. Botstein, A. L. Borresen-Dale & P. O. Brown, 2002. Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc Natl Acad Sci USA,* 99(20), 12963-8.

Potti, A., S. Mukherjee, R. Petersen, H. K. Dressman, A. Bild, J. Koontz, R. Kratzke, M. A. Watson, M. Kelley, G. S. Ginsburg, M. West, D. H. Harpole, Jr. & J. R. Nevins, 2006. A genomic strategy to refine prognosis in early-stage non-small-cell lung cancer. *N Engl J Med,* 355(6), 570-80.

Quint, L. E. & I. R. Francis, 1999. Radiologic staging of lung cancer. *J Thorac Imaging,* 14(4), 235-46.

Raimondi, S., P. Boffetta, S. Anttila, J. Brockmoller, D. Butkiewicz, I. Cascorbi, M. L. Clapper, T. A. Dragani, S. Garte, A. Gsur, G. Haidinger, A. Hirvonen, M. Ingelman-Sundberg, I. Kalina, Q. Lan, V. P. Leoni, L. Le Marchand, S. J. London, M. Neri, A. C. Povey, A. Rannug, E. Reszka, D. Ryberg, A. Risch, M. Romkes, A. Ruano-Ravina, B. Schoket, M. Spinola, H. Sugimura, X. Wu & E. Taioli, 2005. Metabolic gene polymorphisms and lung cancer risk in non-smokers. An update of the GSEC study. *Mutat Res,* 592(1-2), 45-57.

Repace, J. L. & A. H. Lowrey, 1990. Risk assessment methodologies for passive smoking-induced lung cancer. *Risk Anal,* 10(1), 27-37.

Ries, L. A., D. Melbert, M. Krapcho, D. G. Stinchcomb, N. Howlader, M. J. Horner, A. Mariotto, B. A. Miller, E. J. Feuer, S. F. Altekruse, D. R. Lewis, L. Clegg, M. P. Eisner, M. Reichman & B. K. Edwards (eds.), (2008). *SEER Cancer Statistics Review, 1975-2005,* National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/csr/1975_2005/, based on November 2007 SEER data submission, posted to the SEER web site, 2008.

Rosa, P. L., E. Viara, P. Hupe, G. Pierron, S. Liva, P. Neuvial, I. Brito, S. Lair, N. Servant, N. Robine, E. Manie, C. Brennetot, I. Janoueix-Lerosey, V. Raynal, N. Gruel, C. Rouveirol, N. Stransky, M. H. Stern, O. Delattre, A. Aurias, F. Radvanyi & E. Barillot, 2006. VAMP: Visualization and analysis of array-CGH, transcriptome and other molecular profiles. *Bioinformatics,* 22(17), 2066-73.

Rosell, R., F. Cecere, M. Santarpia, N. Reguart & M. Taron, 2006. Predicting the outcome of chemotherapy for lung cancer. *Curr Opin Pharmacol,* 6(4), 323-31.

Ruiz-Godoy, L., P. Rizo Rios, F. Sanchez Cervantes, A. Osornio-Vargas, C. Garcia-Cuellar & A. Meneses Garcia, 2007. Mortality due to lung cancer in Mexico. *Lung Cancer,* 58(2), 184-90.

Sato, M., D. S. Shames, A. F. Gazdar & J. D. Minna, 2007. A translational view of the molecular pathogenesis of lung cancer. *J Thorac Oncol,* 2(4), 327-43.

Schiller, J. H., D. Harrington, C. P. Belani, C. Langer, A. Sandler, J. Krook, J. Zhu & D. H. Johnson, 2002. Comparison of four chemotherapy regimens for advanced non-small-cell lung cancer. *N Engl J Med,* 346(2), 92-8.

Shendure, J., G. J. Porreca, N. B. Reppas, X. Lin, J. P. McCutcheon, A. M. Rosenbaum, M. D. Wang, K. Zhang, R. D. Mitra & G. M. Church, 2005. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science,* 309(5741), 1728-32.

Shepherd, F. A., 1999. Chemotherapy for non-small cell lung cancer: have we reached a new plateau? *Semin Oncol,* 26(1 Suppl 4), 3-11.

Sherbenou, D. W. & B. J. Druker, 2007. Applying the discovery of the Philadelphia chromosome. *J Clin Invest,* 117(8), 2067-74.

Shibata, T., S. Uryu, A. Kokubu, F. Hosoda, M. Ohki, T. Sakiyama, Y. Matsuno, R. Tsuchiya, Y. Kanai, T. Kondo, I. Imoto, J. Inazawa & S. Hirohashi, 2005. Genetic classification of lung adenocarcinoma based on array-based comparative genomic hybridization analysis: its association with clinicopathologic features. *Clin Cancer Res,* 11(17), 6177-85.

Shih, I. M., J. J. Sheu, A. Santillan, K. Nakayama, M. J. Yen, R. E. Bristow, R. Vang, G. Parmigiani, R. J. Kurman, C. G. Trope, B. Davidson & T. L. Wang, 2005. Amplification of a chromatin remodeling gene, Rsf-1/HBXAP, in ovarian carcinoma. *Proc Natl Acad Sci U S A,* 102(39), 14004-9.

Shimamura, T. & G. I. Shapiro, 2008. Heat shock protein 90 inhibition in lung cancer. *J Thorac Oncol,* 3(6 Suppl 2), S152-9.

Sirzen, F., E. Kjellen, S. Sorenson & E. Cavallin-Stahl, 2003. A systematic overview of radiation therapy effects in non-small cell lung cancer. *Acta Oncol,* 42(5-6), 493-515.

Sjoblom, T., S. Jones, L. D. Wood, D. W. Parsons, J. Lin, T. Barber, D. Mandelker, R. J. Leary, J. Ptak, N. Silliman, S. Szabo, P. Buckhaults, C. Farrell, P. Meeh, S. D. Markowitz, J. Willis, D. Dawson, J. K. Willson, A. F. Gazdar, J. Hartigan, L. Wu, C. Liu, G. Parmigiani, B. H. Park, K. E. Bachman, N. Papadopoulos, B. Vogelstein, K. W. Kinzler & V. E. Velculescu, 2006. The Consensus Coding Sequences of Human Breast and Colorectal Cancers. *Science*.

Snijders, A. M., N. Nowak, R. Segraves, S. Blackwood, N. Brown, J. Conroy, G. Hamilton, A. K. Hindle, B. Huey, K. Kimura, S. Law, K. Myambo, J. Palmer, B. Ylstra, J. P. Yue, J. W. Gray, A. N. Jain, D. Pinkel & D. G. Albertson, 2001. Assembly of microarrays for genome-wide measurement of DNA copy number. *Nat Genet,* 29(3), 263-4.

Solinas-Toldo, S., S. Lampel, S. Stilgenbauer, J. Nickolenko, A. Benner, H. Dohner, T. Cremer & P. Lichter, 1997. Matrix-based comparative genomic hybridization:

biochips to screen for genomic imbalances. *Genes Chromosomes Cancer,* 20(4), 399-407.

Strickland-Marmol, L. B., A. Khoor, S. K. Livingston & A. Rojiani, 2007. Utility of tissue-specific transcription factors thyroid transcription factor 1 and Cdx2 in determining the primary site of metastatic adenocarcinomas to the brain. *Arch Pathol Lab Med,* 131(11), 1686-90.

Sun, S., J. H. Schiller & A. F. Gazdar, 2007. Lung cancer in never smokers--a different disease. *Nat Rev Cancer,* 7(10), 778-90.

Takahashi, K., T. Kohno, S. Matsumoto, Y. Nakanishi, Y. Arai, T. Fujiwara, N. Tanaka & J. Yokota, 2007. Clonality and heterogeneity of pulmonary blastoma from the viewpoint of genetic alterations: a case report. *Lung Cancer,* 57(1), 103-8.

Tomida, S., K. Koshikawa, Y. Yatabe, T. Harano, N. Ogura, T. Mitsudomi, M. Some, K. Yanagisawa, T. Takahashi, H. Osada & T. Takahashi, 2004. Gene expression-based, individualized outcome prediction for surgically treated lung cancer patients. *Oncogene,* 23(31), 5360-70.

Tonon, G., K. K. Wong, G. Maulik, C. Brennan, B. Feng, Y. Zhang, D. B. Khatry, A. Protopopov, M. J. You, A. J. Aguirre, E. S. Martin, Z. Yang, H. Ji, L. Chin & R. A. Depinho, 2005. High-resolution genomic profiles of human lung cancer. *Proc Natl Acad Sci U S A,* 102(27), 9625-30.

Travis, W. D., 2002. Pathology of lung cancer. *Clin Chest Med,* 23(1), 65-81, viii.

Tuzun, E., A. J. Sharp, J. A. Bailey, R. Kaul, V. A. Morrison, L. M. Pertz, E. Haugen, H. Hayden, D. Albertson, D. Pinkel, M. V. Olson & E. E. Eichler, 2005. Fine-scale structural variation of the human genome. *Nat Genet,* 37(7), 727-32.

van Wieringen, W. N., J. A. Belien, S. J. Vosse, E. M. Achame & B. Ylstra, 2006. ACE-it: a tool for genome-wide integration of gene dosage and RNA expression data. *Bioinformatics,* 22(15), 1919-20.

Varadhachary, G. R., J. L. Abbruzzese & R. Lenzi, 2004. Diagnostic strategies for unknown primary cancer. *Cancer,* 100(9), 1776-85.

Velculescu, V. E., 2008. Defining the blueprint of the cancer genome. *Carcinogenesis,* 29(6), 1087-91.

Velculescu, V. E., L. Zhang, B. Vogelstein & K. W. Kinzler, 1995. Serial analysis of gene expression. *Science,* 270(5235), 484-7.

Volik, S., B. J. Raphael, G. Huang, M. R. Stratton, G. Bignel, J. Murnane, J. H. Brebner, K. Bajsarowicz, P. L. Paris, Q. Tao, D. Kowbel, A. Lapuk, D. A. Shagin, I. A. Shagina, J. W. Gray, J. F. Cheng, P. J. de Jong, P. Pevzner & C. Collins, 2006. Decoding the fine-scale structure of a breast cancer genome and transcriptome. *Genome Res,* 16(3), 394-404.

Volik, S., S. Zhao, K. Chin, J. H. Brebner, D. R. Herndon, Q. Tao, D. Kowbel, G. Huang, A. Lapuk, W. L. Kuo, G. Magrane, P. De Jong, J. W. Gray & C. Collins, 2003. End-sequence profiling: sequence-based analysis of aberrant genomes. *Proc Natl Acad Sci U S A,* 100(13), 7696-701.

Wain, J. C., S. J. Wilkins, J. R. Benfield & S. S. Smith, 1986. Altered DNA methylation patterns in human lung carcinomas. *Curr Surg,* 43(6), 489-92.

Wang, T. L., C. Maierhofer, M. R. Speicher, C. Lengauer, B. Vogelstein, K. W. Kinzler & V. E. Velculescu, 2002. Digital karyotyping. *Proc Natl Acad Sci U S A,* 99(25), 16156-61.

Weir, B. A., M. S. Woo, G. Getz, S. Perner, L. Ding, R. Beroukhim, W. M. Lin, M. A. Province, A. Kraja, L. A. Johnson, K. Shah, M. Sato, R. K. Thomas, J. A. Barletta, I. B. Borecki, S. Broderick, A. C. Chang, D. Y. Chiang, L. R. Chirieac, J. Cho, Y. Fujii, A. F. Gazdar, T. Giordano, H. Greulich, M. Hanna, B. E. Johnson, M. G. Kris, A. Lash, L. Lin, N. Lindeman, E. R. Mardis, J. D. McPherson, J. D. Minna, M. B. Morgan, M. Nadel, M. B. Orringer, J. R. Osborne, B. Ozenberger, A. H. Ramos, J. Robinson, J. A. Roth, V. Rusch, H. Sasaki, F. Shepherd, C. Sougnez, M. R. Spitz, M. S. Tsao, D. Twomey, R. G. Verhaak, G. M. Weinstock, D. A. Wheeler, W. Winckler, A. Yoshizawa, S. Yu, M. F. Zakowski, Q. Zhang, D. G. Beer, Wistuba, II, M. A. Watson, L. A. Garraway, M. Ladanyi, W. D. Travis, W. Pao, M. A. Rubin, S. B. Gabriel, R. A. Gibbs, H. E. Varmus, R. K. Wilson, E. S. Lander & M. Meyerson, 2007. Characterizing the cancer genome in lung adenocarcinoma. *Nature,* 450(7171), 893-8.

Wilson, I. M., J. J. Davies, E. Vucic, W. W. Lockwood, C. MacAulay, S. Lam & W. L. Lam, 2006. Methylome profiling identifies hypomethylation and amplification in lung cancer genes modulating p53 degradation. *In preparation,* 1(1), 1-10.

Winton, T., R. Livingston, D. Johnson, J. Rigas, M. Johnston, C. Butts, Y. Cormier, G. Goss, R. Inculet, E. Vallieres, W. Fry, D. Bethune, J. Ayoub, K. Ding, L. Seymour, B. Graham, M. S. Tsao, D. Gandara, K. Kesler, T. Demmy & F. Shepherd, 2005. Vinorelbine plus cisplatin vs. observation in resected non-small-cell lung cancer. *N Engl J Med,* 352(25), 2589-97.

Workman, P., 2004. Altered states: selectively drugging the Hsp90 cancer chaperone. *Trends Mol Med,* 10(2), 47-51.

Yung, R. C., 2003. Tissue diagnosis of suspected lung cancer: selecting between bronchoscopy, transthoracic needle aspiration, and resectional biopsy. *Respir Care Clin N Am,* 9(1), 51-76.

Zhao, X., C. Li, J. G. Paez, K. Chin, P. A. Janne, T. H. Chen, L. Girard, J. Minna, D. Christiani, C. Leo, J. W. Gray, W. R. Sellers & M. Meyerson, 2004. An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays. *Cancer Res,* 64(9), 3060-71.

Zhu, C. Q., J. C. Cutz, N. Liu, D. Lau, F. A. Shepherd, J. A. Squire & M. S. Tsao, 2006. Amplification of telomerase (hTERT) gene is a poor prognostic marker in non-small-cell lung cancer. *Br J Cancer,* 94(10), 1452-9.

2. KEY FEATURES OF BAC ARRAY PRODUCTION AND USAGE AND ACCRUAL OF CLINICAL LUNG TUMOUR TISSUE[1]

2.1. Introduction to BAC array production

Somatic DNA copy number alterations are hallmarks of cancer, disrupting the expression of oncogenes and/or tumor suppressor genes, while constitutional DNA copy number variations have been associated with developmental disorders (de Vries et al. 2005; Hanahan & Weinberg 2000). Identification of these alterations will impact disease susceptibility characterization, disease subclassification, and treatment suitability. It will also lead to the development of novel prognostic/diagnostic markers and provide new therapeutic targets.

2.1.1. Surveys of DNA copy number changes

Comprehensive analysis of genetic alterations requires high resolution techniques (Garnis et al. 2004a; Davies et al. 2005; Lockwood & Chari et al. 2005). Initial array CGH experiments were performed on cDNA microarrays (Pollack et al. 1999). However, cDNA targets lack introns that are present in the genomic probe, thus resulting in low signal-to-noise ratios[2]. Oligonucleotide-based platforms, such as those used for single-nucleotide polymorphism (SNP) analysis and representative oligonucleotide microarray analysis (ROMA), are powerful means of assessing DNA copy number integrity (Lucito et al. 2003; Matsuzaki et al. 2004). However, even with >100,000 loci represented on SNP arrays, only a subset of loci will be informative. Also, there is evidence that the use of genomic reduction techniques (e.g. whole-genome sampling) and PCR amplification steps that are typically used for oligo-array analysis contribute to experimental variability, bias, and loss of feature details (Davies et al. 2005; Bignell et al. 2004).

---

[1] A version of this chapter has been published: Buys TPH, Wilson IM, Coe BP, Lockwood WW, Davies JJ, Chari R, DeLeeuw RJ, Shadeo A, MacAulay C, Lam WL. (**2007**) "Key Features of BAC Array Production and Usage" in *DNA Microarrays (Methods Express Series)* (Schena M, ed.), Scion Publishing, Ltd. Bloxham, Oxfordshire, UK, pp.115-145 (ISBN: 9781904842156).

[2] In this chapter, the term "target" refers to elements displayed on the array, while the term "probe" refers to the labeled sample applied to the array.

Arrays comprised of bacterial artificial chromosomes (BACs) offer another means of high throughput DNA copy number analysis. These large insert clone arrays can be obtained from several sources (see Table 2.1 for useful web links). BAC arrays have lower DNA sample input requirements than other platforms, require no amplification steps, and are more sensitive at detecting single-copy changes than other platforms. While arrays with smaller probe targets will ultimately provide higher resolutions, current BAC array platforms combine high resolution with high sensitivity to give the most reliable analysis.

2.1.2. BAC arrays

BAC arrays can be subdivided into two main categories: disease-specific (or region-specific) and genome-wide. These categories can be further divided into low and high resolution depending on the number of clones used to span a given region.

2.1.2.1. Disease/region-specific BAC arrays

A number of arrays exist for examining specific diseases or specific chromosomal regions (Greshock et al. 2004; Kohlhammer et al. 2004; Massion et al. 2002; Mantripragada et al. 2003; Nessling et al. 2005; Roerig et al. 2005; Cheung et al. 2005; Garnis et al. 2004b; Schwaenen et al. 2004; Coe et al. 2005; Garnis et al. 2003; Garnis et al. 2005a; Garnis et al. 2004c; Garnis et al. 2005c; Garnis et al. 2004d; Henderson et al. 2005; Kameoka et al. 2004; van Duin et al. 2005; Buckley et al. 2002; Davison et al. 2005; Zafarana et al. 2003; Redon et al. 2005; Solomon et al. 2004). Manufacturers of commercially available platforms are listed in Table 2.1. All these platforms offer reliable interrogation of copy number for their respective targets and their focused design have proven useful in both cancer and constitutional disease studies. However, they are of limited use to researchers investigating novel loci. For effective design, selection of regions for these arrays requires *a priori* knowledge of disease-driving alterations.

2.1.2.2. Whole genome sampling and complete coverage BAC arrays

Gene discovery research is best served by the use of CGH arrays that provide unbiased coverage of the entire human genome. These include interval marker-based[3] arrays such as the Spectral Genomics Spectral Chip 2600 and the UCSF HumArray which cover the human genome at a density of ~1 clone per Mb ([Snijders et al. 2001]; Table 2.1). Due to the relationship between genome coverage and the probability of alteration detection, marker-based methodologies will not typically uncover changes smaller than their average clone spacing, meaning that small alterations may not be detected and alteration boundaries may not be fine-mapped (Davies et al. 2005). Additionally, samples of poor quality, such as those extracted from paraffin embedded tissue, tend to increase the noise observed in experiments, thus further reducing the resolution (Garnis & Coe et al. 2005b). To detect small scale genomic alterations, the sub-megabase resolution tiling-set (SMRT) CGH array was developed (Ishkanian et al. 2004). This array is comprised of >32,000 overlapping BAC clones, meaning that there is no need to infer alteration status between clones. SMRT array CGH enables the fine mapping of breakpoints to within a single BAC clone and the detection of alterations as small as 40-80 kb.

2.1.3. Platform choice

Ultimately, available arrays may not meet the precise needs of a researcher, leaving labs to produce their own array platforms. What follows is a description of the steps required for BAC array production and experimental use, including specific details based on our experience manufacturing a high density SMRT array. We will provide protocols and troubleshooting for array production; sample preparation methods; labeling, hybridization, and scanning of hybridized arrays; and analysis of experimental data.

[3] The term “marker-based” refers to those platforms consisting of targets that sample the genome at various intervals.

2.2. Manufacturing BAC arrays

2.2.1. Preparation of BAC clones for spotting

2.2.1.1. Description of BAC clones and available libraries

Due to ease of use and the ability to carry large inserts (50-200 kb), the majority of the BAC libraries constructed to date have used the modular BAC vector pBACe3.6 (Frengen et al. 1999). Several human BAC clone libraries are currently available, the most commonly referenced include: Roswell Park Cancer Institute (RPCI) and Caltech (CTD) ([Osoegawa et al. 2001]; Table 2.1). The SMRT array is composed of 32,433 BAC clones selected from the RPCI-11, RPCI-13, and CTD-D libraries to produce a 1.5 times overlapping coverage of the human genome (Ishkanian et al. 2004; Krzywinski et al. 2004).

2.2.1.2. Validation of BAC clone identity

Clone identity can be verified through multiple means. *Hind*III fingerprints of BAC clones can be compared to the physical map of the human genome using Fingerprint Contig (FPC) BAC fingerprint database (Marra et al. 1997; McPherson et al. 2001). Banding patterns will be unique to the clone while a compilation of digests from a series of clones will reveal the relationship between the clones with respect to chromosomal position and thus is a less expensive and less time consuming alternative to other approaches. At the time of our SMRT array construction, we used this approach because the draft sequence of the human genome was not finalized and fingerprint contigs could be used to bridge sequence gaps (Krzywinski et al. 2004). Other means of confirming amplified fragment pool (AFP) clone identity include fluorescence *in situ* hybridization (FISH) and sequencing. High throughput AFP sequencing as described by Watson *et al*. is more efficient for large clone sets such as the one used for the SMRT array (Watson et al. 2004).

2.2.1.3. Amplification of BACs

Because BACs are low copy vectors, DNA yield per cell is low. Additionally, DNA isolation from primary cells is not a high throughput solution to generating a replenishable stock of spotting solution where best performance requires 500-1000

ng/µl. Due to the large clone set of the SMRT array and the efficient use of resources in amplification, we have adopted a PCR based strategy to generate spotting solution from high throughput/low yield BAC DNA isolations. The two most common methods of BAC clone amplification include degenerate oligonucleotide primer PCR (DOP-PCR) and linker mediated PCR (LM-PCR). We elected to use LM-PCR because it offers a linear, unbiased amplification as linkers with primer sites are ligated to all fragments generated by digestion using a four-cutter restriction enzyme (Pfeifer et al. 1989; Telenius et al. 1992; Fiegler et al. 2003). An LM-PCR protocol used for SMRT array production is provided ((Watson et al. 2004), Figure 2.1, Protocol 1). A major benefit to this method is the generation of source LM-PCR #1 product which can be stored and used to generate additional LM-PCR #2 product as needed. Both DOP-PCR and LM-PCR can be employed to modify fragments promoting their interaction with different slide surface chemistries (for more on this topic, see section 2.2.2.1).

2.2.1.4. Spotting solution

A variety of spotting solutions are available, including Micro Spotting Plus (MSP) (Telechem), Pronto (Corning), Sodium Phosphate, and 3X saline sodium citrate (SSC). Our experiences with MSP and SSC-based buffers have been positive. One important consideration is that evaporation is a problem with all spotting solutions. Non-proprietary solutions (i.e. those with known reagents) are advantageous because the experimenter can reconstitute them by replenishing the appropriate solvents.

2.2.2. Printing a BAC array

After determining the number and desired density of the spotted targets, users must choose an array printer, printing pins, a type of slide chemistry, and a PCR protocol for target DNA.

2.2.2.1. Slide surface chemistry

There are a variety of slide types available for spotting BAC microarrays. Two common slide types are amino-silane (which bind DNA by charge interactions) and aldehyde (which covalently bind to 5’ amino-modified DNA). While aldehyde-coated slides would not be appropriate for binding raw BAC DNA, they are ideal for PCR products, as each

primer can have an amino-modified 5' terminus. Other characteristics that are important in the selection of microarray slides are the hydrophobicity, uniformity of chemical coating, and the optical purity of the glass used. Hydrophobicity and uniformity of chemical coating are of utmost importance due to the effects of these parameters on spot size and morphology. Glass with high optical purity (i.e. low intrinsic autofluorescence) is critical in reducing background signal.

A major consideration when selecting a microarray slide type is the ability to block or inactivate the unspotted portions of the slide to reduce background signal. While amino-silane and similar chemistries generally require a pre-hybridization to block the still-active slide chemistry, aldehyde-coated slides can be inactivated after being spotted (see section 2.2.4).

We currently use aldehyde-coated slides for production of the SMRT array because of the covalent binding of spotted 5' amino-modified DNA, the high hydrophobicity, and the ability to inactivate unspotted regions of the slide (providing the best overall signal to noise ratios). We find that coating is variable amongst manufacturers and we recommend that each batch be carefully tested prior to full scale production (see troubleshooting, section 2.7).

2.2.3. Array printers

For the production of arrays from PCR products, we have found that machines capable of using quill or solid style printing pins perform best (see section 2.2.3.5). Many manufacturers provide high quality printers which differ in throughput, accuracy, sample handling, and ease of use. When deciding on the printer that best suits your needs, there are many points to consider.

2.2.3.1. Plate handling features

The ability to handle multiple plates through a stacking mechanism is essential to reducing the labor involved in long spot runs. Additionally humidifying the plate enclosure is important in preventing evaporation of the valuable printing solution (more on humidity in 2.3.3).

2.2.3.2. Printer throughput

Depending on your application, you may require a high throughput printer capable of holding hundreds of slides or a smaller system which handles fewer slides. Array printer pricing increases directly with platen[4] size (which determines the number of slides spotted per session), however the labor cost per slide increases with smaller capacity array printers. Thus it is critical to obtain an array printer which can meet your needs both at the time of purchase and for future expansion. The speed of printing is affected by several factors, with pin wash station efficiency and the number of pins in the print head having the largest impact on total printing time. For most high density arrays, a 48 pin print head represents the most efficient option for printing from 384 well plates.

2.2.3.3. Air filtration and humidity control

Humidity is essential to preventing evaporation of spotting material from the printing plate and controlling printed spot morphology. Control of air purity is also important for high density arrays because of the small size of printed features (~100 μm) and the ability for small dust particles to clog the quill pins (Hegde et al. 2000; McQuain et al. 2003). For this reason, High Efficiency Particulate Air (HEPA) filters are common on printers.

2.2.3.4. Levelness of platen

The amount of time a pin spends in contact with a slide as well as the impact speed can have a drastic effect on feature size. The best printing results can be obtained when the pins just touch the top of the slide. In this situation a 50 μm difference in depth can cause a noticeable difference in print quality. Unfortunately, many printers can exhibit variations in excess of 100 μm across the surface of the bed. While calibration may be able to compensate for this, it is critical to check with the manufacturer as to their precise specifications and to assay for variability with test prints.

[4] The term “platen” refers to the slide bed.

2.2.3.5. Printing pins

The choice of a printing pin is critical to the size of spots you wish to produce and this in turn has a direct effect on the maximum density at which you can print. Higher printing densities allow more clones to be placed onto a single microarray slide or the ability to reduce the surface area of the printed array. This allows the use of lower volumes of hybridization solution and the ability to print multiple arrays on a single glass slide.

There are two major styles of pins suitable for the production of BAC microarrays. Solid pins are simple highly durable pins which produce one spot per dip into the printing solution. Quill style pins contain a slit in the tip of the pin that acts as a spotting fluid reservoir drawing up liquid from the printing plate by capillary action. Quill pins are capable of producing many (~100-200) spots per dip into the printing solution, but at the cost of durability. For any high throughput or high density printing, quill pins are the best option as they substantially reduce printing times and overall cost.

The second major factor to consider is the diameter of the pins end. All pins produce spots slightly larger than the pin end. This effect is strongest in quill style pins, although smaller pins should usually produce smaller spots. Smaller pins offer the benefit of high density printing but sacrifice the durability of pins with larger end diameters. For pins capable of the highest printing densities, the impact speed needs to be reduced in order to preserve pin life and minimize spot size.

All printing pins require careful attention to preserve their performance. Even in array printers with high performance pin cleaning stations (including sonicators or multiple wash solutions), residue will build up. This causes the pins to eventually demonstrate reduced print quality or lose their ability to print altogether, resulting in missing grids that sometimes emerge during print runs. Thus, it is important to follow the manufacturers' recommended cleaning protocols between every large print run or even during a long run if quality is observed to decrease.

We have used two major brands of pins with good results. We find that Telechem produces high quality pins that allow the printing of small features (90 μm spot diameter with optimization), producing consistent high density arrays. Genetix brand pins are constructed of a more durable stainless steel and perform well for the production of feature sizes on the order of 100 μm. However, it is difficult to acquire pins which consistently match the Telechem SMP pins in feature size ([Hegde et al. 2000; McQuain et al. 2003], Table 2.1).

2.2.3.6. Plate handling

The handling and choice of spotting solution source plates is of critical importance to getting the most from your solution. It is important to choose a plate which allows low volumes of spotting solution (ideally 10 μl), resists warping, and is rigid enough to maintain the tight tolerances required by the array printer. Integrated lids are critical to preventing evaporation while plates are present in the plate stacker. The use of foil sealing tape to seal plates immediately after use and subsequent freezing of plates will prevent evaporation and degradation, increasing the number of prints attainable from a single preparation of solution. Plate choice is typically linked to the array printer used, though other options may be possible under specific circumstances.

Although most modern array printers include an integrated plate stacker for handling a large set of plates, it is important to reduce the amount of time each plate stays thawed and unsealed. In our experience, limited evaporation is observed if plates are left unsealed with lids on for up to 1.5 hours at 50% humidity. If the plate stacker offers independent humidity control, a higher humidity setting can be used to prevent solution evaporation without impacting spot morphology.

2.2.3.7. Optimizing printing parameters

After acquiring an array printer, preparing printing solution, and selecting a slide type, there are several optimizations that need to be performed to produce consistently high quality arrays. It is well known that humidity has a drastic effect on spot morphology. We find that a setting of 50-60% humidity produces the best results. Additionally, the speed at which a pin hits the slide surface and the time it dwells on the surface (affected

by the pin overstroke) are important considerations. Ultimately, it is important to realize that there are no fixed settings which will work in every environment, with all equipment, and for all printing solutions. Thus it is of vital importance that each parameter is optimized through a series of test prints (McQuain et al. 2003).

2.2.4. Processing

The goal of processing slides after spotting is twofold: to remove any unbound material and to inactivate reactive groups not used to bind BAC DNA. Removing unbound material is essential because this material can move under the coverslip during hybridization and cause both poor spot morphology and high background signal. Inactivation of reactive groups can help to reduce background signal strength. The SMRT array is typically spotted on aldehyde slides and for this surface chemistry, inactivating reactive groups with a strong reducing agent is an effective means of reducing background (e.g. sodium borohydride, $NaBH_4$).

2.3. Selection of samples for use on the BAC array

2.3.1. Samples quality and quantity

The SMRT array is currently the most high throughput and comprehensive means of profiling genomic alterations in archival samples[5]. This is because of its minimal input DNA requirements and its ability to withstand non-target cell DNA contamination (Garnis & Coe et al. 2005b; Davies et al. 2005). Currently, DNA from archived specimens cannot be applied to other high throughput platforms such as the Affymetrix 100k SNP array (Conrad 2005). Ultimately, the main issues to take into account when assessing applicability of a sample for a BAC array experiment are the relative heterogeneity of the contributing cell population, the quantity of DNA, and the quality of DNA.

If tissue heterogeneity is a concern, manual or laser-assisted microdissection allows targeting of only desired cells within a tissue cross-section (Emmert-Buck et al. 1996; Bonner et al. 1997; Rekhter & Chen 2001). Even with an increased ability to tolerate

[5] "Archival samples" are typically formalin fixed and paraffin embedded (FFPE). These samples are stored in hospital archives.

heterogeneity, selection of a subpopulation of cells will produce clearer genomic profiles and facilitate more effective analysis (Garnis & Coe et al. 2005b).

While a number of other platforms employ genome sampling and amplification steps, the low DNA input requirement of BAC arrays generally removes this need (Davies et al. 2005). Currently, different BAC array groups report DNA input requirements in the range of 300 ng to 3 µg (Pinkel & Albertson 2005b), though our experience with the SMRT array is that 50-400 ng of DNA per slide will suffice. Eliminating the requirement of an amplification step is advantageous as factors such as template length, secondary structure, and GC content can contribute to DNA segment misrepresentations that bias analysis (Hughes et al. 2005; Lasken & Egholm 2003).

DNA samples derived from fresh, frozen, and archival specimens can be analyzed by BAC array CGH. Relative to formalin fixation (which can degrade nucleic acids), freezing a sample better preserves DNA quality and allows concurrent isolation of RNA and proteins to assess gene expression. However, for clinical samples, fixed tissues produce more effective histological references, facilitating the selection of a desired subpopulation of cells in a heterogeneous sample. In addition, archived samples may be associated with a wealth of clinical data (e.g. outcome), making them extremely valuable in retrospective analyses. In our experience, regardless of the source material, the DNA extraction protocol can also significantly impact DNA quality and yield. For more factors affecting DNA sample quality, refer to sections 2.7 and 2.8.

2.3.2. Choice of reference sample

In terms of reference samples for competitive hybridization, depending on the experiment, different samples will satisfy user needs (see Figure 2.2). Issues that need to be considered in selecting the reference sample are sex (matched vs. mis-matched to the test sample), composition (individual vs. pooled from multiple sources), and source (allogenic vs. autogenic). A single reference type should be used for an entire experimental set. Because reference DNA will contain copy number variations (CNVs) (i.e. natural polymorphisms), these need to be taken into account during experimental analysis. Significantly, further characterization of CNVs using a tiling-set array will build

on existing work and serve as an invaluable baseline to identify population variance as well as susceptibility loci driving various constitutional diseases (Iafrate et al. 2004; Sebat et al. 2004; Sharp et al. 2005; de Vries et al. 2005; Wong et al. 2007).

2.4. BAC array hybridization

2.4.1. Approaches to probe generation and labeling

There are numerous methods to differentially label sample and reference DNA with fluorescent nucleotides (such as cyanine-3 and cyanine-5 dCTPs) for use in BAC array CGH. The ability to accurately detect single copy chromosomal gains and losses depends greatly on the ability to generate high quality labeled probes (Lieu et al. 2005; Tsubosa et al. 2005). Thus, many factors influencing labeling, such as the incorporation efficiency, the spectral separation of fluorophores, and the non-biased amplification of DNA, must be considered to obtain maximum sensitivity and accuracy in detecting copy number alterations.

2.4.1.1. Sample handling prior to probe generation

The two main approaches to DNA labeling are whole genome and representational (see Figure 2.3). Whole genome labeling is the primary method used for BAC array CGH as it creates probe from all of the initial material (see section 2.4.1.2). This method leads to the linear amplification and labeling of the entire genomic DNA sample without complexity reduction. Genome representation, also known as “genome sampling”, is a complexity reduction approach used to reduce cross hybridization noise in oligonucleotide platforms (Lisitsyn & Wigler 1993). This is not needed in BAC array CGH (Lucito et al. 2000; Bignell et al. 2004; Davies et al. 2005). In contrast to whole genome labeling, the representational approach enriches for short DNA fragments by performing a restriction enzyme digest followed by linker-mediated PCR amplification.

2.4.1.2. Probe generation

The most common whole genome probe generation approach is random primer DNA labeling, originally developed by Feinberg and Vogelstein (Feinberg & Vogelstein 1983; Feinberg & Vogelstein 1984; Lieu et al. 2005; Tsubosa et al. 2005). This procedure

facilitates integration of cyanine-dNTP nucleotide analogues[6] into template DNA and requires between 50 ng and 3 µg of material per reaction ([Pinkel & Albertson 2005b], unpublished data). Random priming uses a mixture of all possible sequences of short primers (usually hexamers or octamers) that hybridize to the template as starting points for DNA synthesis (see Figure 2.3). This is done to ensure an equal degree of probe generation from the entire length of the template DNA. The Klenow enzyme, the large fragment of DNA polymerase I, is then used to synthesize the complementary strand from the 3' OH of the primer, incorporating labeled nucleotides into the complementary strand. Because it has the capacity to displace strands on the template DNA, fragments can be generated that are larger than the spaces between primers. Random priming leads to at least a four-fold non-biased linear amplification of the starting material so that larger amounts of probe are used in the ensuing hybridization, an attractive option when using low DNA yield clinical samples (Lieu et al. 2005; Pinkel & Albertson 2005a).

Whole genome methods employed to label DNA for array CGH experiments can be direct or indirect (see Figure 2.3) (Richter et al. 2002; Yu et al. 2002; Xiang et al. 2002; Solinas-Toldo et al. 1997). Indirect labeling requires secondary dye coupling steps and is not widely used in array CGH experiments at this time. With direct labeling, tagged nucleotides are directly incorporated into probe generated from template DNA, simplifying the labeling procedure.

Troubleshooting for labeling experiments can be found at the end of the chapter. Please note that spectrophotometer readings only measure the amount of cyanine-labeled dNTPs incorporated into the probe and not the emission of the dyes, therefore fluorometric readings may more accurately reflect the relative activity of the dyes (Yu et al. 1994). Also, please note that a major concern when using cyanine dyes is their sensitivity to environmental agents. Extended exposure to light or high levels of atmospheric ozone have been shown to affect the fluorescence of these dyes and should be limited during array experiments (Petrescu et al. 2003; Fare et al. 2003).

[6] Cyanine dyes have proven useful due to their detection sensitivity and efficiency as a polymerase substrate.

2.4.2. C_0t-1 DNA

Whenever using probes or targets that are created from genomic DNA, blocking repetitive sequences is required to achieve sequence-specific fluorescent ratios. This is typically achieved through the use of C_0t-1 DNA (Britten & Kohne 1968; Marx et al. 1976; Schrock et al. 1996). C_0t-1 is species-specific genomic DNA that has been enriched for repetitive sequences through sonication and controlled denaturation/re-annealing. The binding of the C_0t-1 DNA to the repetitive elements present in the generated probe prevents them from binding to targets spotted onto the array. It is typically used in excess of the probe DNA to ensure adequate blocking, with the final concentration ranging from 1-20 µg/µl in 15-110 µl of hybridization buffer and a ratio to probe DNA of approximately 3:1 (Carter et al. 2002). Labeled probes are precipitated with C_0t-1, redissolved in hybridization buffer, and then typically blocked prior to hybridization. In our experience, C_0t-1 DNA varies considerably from manufacturer to manufacturer, and even from lot to lot. We strongly recommend that test array experiments and rigorous quality control standards be applied prior to purchasing any large amount of C_0t-1 DNA (see troubleshooting, section 2.7).

2.4.3. Hybridization buffer

Hybridization buffers are designed to lower the melting point of DNA and maintain stringency, enabling hybridization at lower temperatures. Formamide or urea may be used to lower the melting point of DNA, with formamide being more common and generally used at a 50% concentration (Casey & Davidson 1977). However, urea-based hybridization solutions are desirable as they are less toxic than their formamide counterparts (Simard et al. 2001). Every increase of 1% in formamide concentration lowers the melting point of DNA by 0.7°C (Casey & Davidson 1977). In the presence of 10% dextran sulfate or polyethylene glycol, the hybridization rate is increased tenfold (Wahl et al. 1979; Renz & Kurz 1984; Amasino 1986), as the effective concentration of the probe is increased. 2X SSC is common in array CGH hybridization buffers (Carter et al. 2002). Detergents are also used frequently, with sodium dodecyl sulfate (SDS) being most common. For exact concentrations used by different groups please consult

(Carter et al. 2002). Commercially obtained hybridization buffers may offer a higher level of consistency; we have had success with DIG Easy from Roche (Table 2.1).

Addition of other macromolecules such as yeast tRNA and sheared herring sperm DNA to a hybridization solution is also a common practice. This is done in an effort to minimize non-specific binding to both the slide and the target DNA sequences. Yeast tRNA is used by various groups at final concentrations ranging from 50 ng/µl to 10 µg/µl and sheared herring sperm at final concentrations ranging from 50 ng/µl to 400 ng/µl in hybridization volumes of 15-110 µl (Carter et al. 2002).

2.4.4. Hybridization

In addition to buffer and temperature, there are other factors to consider before setting up a hybridization experiment. These include whether or not a pre-hybridization is necessary, the optimal buffer volume and probe concentration, and the type of hybridization performed.

Pre-hybridization can be used to block repeat DNA in array targets, reducing non-specific binding of probe DNA (Southern 1975). In addition, pre-hybridization will block reactive groups on slides which were not chemically inactivated during pre-processing steps (e.g. amine-coated slides – see section 2.2.2.1). For aldehyde slides that are chemically inactivated, this step is not necessary as pretreatment with $NaBH_4$ reduces reactive surface into alcohol groups that are unable to bind to DNA.

All that is required to perform a hybridization experiment is contact between the array and the generated probe. Performing hybridizations under slide coverslips allows for the use of smaller probe concentrations and smaller volumes, an important feature when test material is limited (e.g. most clinical specimens) (see Figure 2.4). One concern with this approach is the lack of probe diffusion, limiting the chance of the complementary probe strands coming into contact with spotted targets on the slide surface (Borden et al. 2005). Automated hybridization apparatus that use air bladders or rocking to ensure greater probe diffusion reportedly give increased sensitivity, though they ultimately require more probe in order to work effectively, meaning that they are

best used for experiments where test DNA is not limiting (Adey et al. 2002; Carter et al. 2002).

"Sandwich hybridizations" offer one way to control for hybridization differences between arrays of a multi-slide set (Ting et al. 2003). Briefly, this entails placing one array on top of the other with the spotted surfaces facing each other. The benefit of this approach is that both slides are exposed to identical experimental conditions. However, drawbacks to this technique include the fact that arrays seem to be more prone to drying out and that some hybridization chambers are not deep enough to accommodate stacked slides.

2.4.5. Washing hybridized slides

Once the hybridization is complete, the slides must be washed to remove non-specifically bound probes (see Protocol 3, section 2.6.3). Several solutions are commonly used for this purpose, typically including formamide, SSC, and SDS (Carter et al. 2002). Cyanine dyes have been shown to be particularly sensitive to atmospheric ozone degradation during the washing and subsequent drying steps (Fare et al. 2003). Thus, ozone levels should be controlled to obtain best quality array CGH results (see troubleshooting, section 2.7).

2.5. Post-hybridization BAC array scanning and experimental analysis

2.5.1. Scanning hybridized slides

Post-hybridization image acquisition is a critical step in the production of array CGH data. With BAC array CGH, slides rarely exhibit the high dynamic range requirements of expression microarrays. This is due to the relatively low ratios observed for single copy changes which account for the greatest number of alterations in the average cancer sample (Hyman et al. 2002). Additionally, high probe complexity and C_0t-1 blocking can result in array slides which exhibit much lower peak fluorescence than a typical expression microarray. Due to these traits it is much more important to obtain a scanner capable of low noise acquisition of low intensity signals than to be concerned about the highest dynamic range.

The two primary scanner technologies available today incorporate a laser/photo-multiplier tube (PMT) imager or a charge-coupled device (CCD) sensor in combination with a white light source and excitation/emission filters (Burgess 2001). In our experience, current CCD-based systems offer the best low intensity performance, however the need to frequently replace the excitation light source and the slow scan times limit their effectiveness in a high throughput setting. Laser-based scanners are the more traditional microarray imaging systems and can vary drastically in performance.

All scanners require adjustment of the imaging sensors and excitation sources to achieve an optimal image which lies in the scanners linear dynamic range. CCD-based scanners are arguably the easiest to pre-configure as they exhibit very wide linear ranges and all configuration settings behave in a linear manner (i.e. doubling the exposure time results in a doubled intensity). Laser-based scanners require a more complex adjustment where both the laser power and PMT sensitivity may be adjusted and neither represents a linear scale. This problem however is countered by the automatic adjustment algorithms incorporated into many newer scanners which allow true automated image acquisition.

2.5.2. Analysis of BAC hybridization results

After the image is scanned, there are a series of steps that need to be performed before a sample's DNA copy number can be assessed. These include image processing, data filtering and visualization, and statistical analysis. Furthermore, prior to visualization and analysis, normalization of the data may also be required (see troubleshooting, section 2.7).

2.5.2.1. Image processing and data filtering

Upon scanning of the image, signal intensity from each spot on the array must be quantified. Specifically, the intensity values of the two fluorophores need to be calculated and, depending on which dyes are coupled with which sample, the appropriate $\log_2$ ratio [test sample intensity]/[reference sample intensity] needs to be calculated. There are many programs available that can perform image quantification,

some of which are bundled with the scanners and others that are available as stand-alone applications (Table 2.1).

After the image has been processed and ratios have been calculated, one can inspect the image and values and find that there are a small percentage of spots of suboptimal quality. Moreover, with any DNA microarray there will be a small percentage of spots that will not contain usable information. Abnormal shape morphology, low intensity values, low signal-to-noise ratio (SNR), and contamination (e.g. dust particles) are potential factors leading to exclusion of such suboptimal spots.

2.5.2.2. Normalization

Normalization addresses broader experimental trends that cannot be resolved through simple spot exclusion. It is required to account for the various CGH biases that produce suboptimal results. Differential dye effects (e.g. cyanine-3 and cyanine-5 detection or incorporation differences), variation in scanning parameters between experiments, spatial effects associated with spot locations, slide intensity gradients, and differences in the amount of starting material between the test and reference sample are among the factors that normalization addresses.

There are two broad levels of normalization that can be performed: global or local. Most methods of normalization can be used in either a global or local context. Two of the most common methods are scaling normalization and locally weighted scatter-plot smoothing (LOWESS). Scaling normalization is usually performed by transforming each of the individual intensities such that, for example, the mean or median intensity for each channel in each array or over the whole set of arrays is identical across all experiments. LOWESS is a normalization approach which uses a locally-based least-squares-weighted regression model to fit subsets of data until all data points have been evaluated under the regression function (Cleveland 1979; Quackenbush 2002). It is a locally-based method because evaluation of a particular data point uses its neighboring points as well.

Many successful normalization approaches for expression microarray data are currently being applied to DNA microarrays. However, a normalization framework was recently developed in the context of DNA BAC arrays which targets three areas of potential bias: spot position, clone origin, and intensity data (Figure 2.5) (Khojasteh et al. 2005). As can be seen in the post-normalization SeeGH karyogram, programs such as this greatly reduce the systematic biases in the process which are usually perceived as noise. This facilitates clearer assessment of DNA copy number. Continuing development of normalization methods for DNA copy number data will improve array CGH data analysis.

2.5.2.3. Visualization and statistical analysis

After the elimination of poor quality spots and transformation of the data represented in the higher quality spots, the final step is to assess statistical significance in the computed ratios and visually represent the data to highlight DNA copy number alterations. Currently, there are freely available software tools that employ different techniques to assess statistical significance (Lockwood & Chari et al. 2005; Lai et al. 2005). Software choice is ultimately made based on user needs. Whereas clinical usage requires a limited robust form of software that allows only selected types of data interrogation, exploratory scientific research requires a more sophisticated and comprehensive application (Lockwood & Chari et al. 2005; Chi et al. 2004; Margolin et al. 2005). *SeeGH* software is currently the only program that properly displays tiled clone data ((Chi et al. 2004); Table 2.1). Future applications will enhance the power of tiled clone platforms by providing analysis approaches that capitalize on the degree of overlap and the polymorphic status for the arrayed elements.

2.6. Detailed protocols

2.6.1. Protocol 1 – LM-PCR of BAC clones

Equipment and reagents

- BAC DNA
- Taq polymerase + buffer (5 U/µl; Promega)

- T4 DNA Ligase (400 U/µl; New England Biolabs)
- *Mse*I (10 U/µl; New England Biolabs)
- *Mse*I Buffer (10X; New England Biolabs)
- dNTPs (10mM; Promega)
- $MgCl_2$ (25mM; Promega)
- *Mse*I-long oligonucleotide (10 µM; Alpha DNA; 5'-AGTGGGATTCCGCATGCTAGT-3')
- *Mse*I-short oligonucleotide (10 µM; Alpha DNA; 5'-TAACTAGCATGC-3')
- sodium acetate (3.0 M, pH 5.5)
- 20X SSC Buffer
- 96 well plates
- Incubator (16°C)
- Thermocycler
- absolute ethanol
- centrifuge

Method

1. Transfer 50 ng of BAC DNA from each clone to a 96 well plate.
2. Add 5 Units *Mse*I restriction enzyme and appropriate buffer. (Dilute to final reaction volume of 40 µl with dH_2O.)
3. Digest for 8 h at 37°C then heat inactivate the *Mse*I by incubating at 65°C for 10 min.
4. Transfer 4.0 µl of digestion product to a new 96 well plate.
5. Linker reaction:
   - 4.0 µl DNA (from Protocol 1, Step 4).
   - Combine 0.8 µl of *Mse*I-long and 0.8µl *Mse*I-short primers and pre-anneal at room temperature for 5 min.
   - Add 80 CEL (Cohesive End Ligation) Units T4 DNA ligase and appropriate buffer. (Dilute to final reaction volume of 40 µl with dH_2O.)
6. Incubate mixture for 12-16 h at 16°C.

7. Remove 2.5 µl of the ligation product to use in 50 µl PCR reaction (termed "LM-PCR #1").

   LM-PCR #1 reaction:

   - 2.5 µl linker-ligated DNA
   - 16 µl $MgCl_2$
   - 5 µl each dNTP
   - 2 µl *Mse*I long primer (amino-modified if using aldehyde slides)
   - 5 Units Taq polymerase and appropriate buffer. (Dilute to final reaction volume of 50 µl with dH_2O.)

   LM-PCR #1 cycling parameters:

   a. 3 min, 95°C

   b. 1 min, 95°C
   c. 1 min, 55°C
   d. 3 min, 72°C
   } 30 cycles

   e. 10 min, 72°C

8. Remove 0.25 µl of LM-PCR #1 to use in another 50 µl PCR reaction (termed "LM-PCR #2")

   LM-PCR #2 reaction:

   - As for LM-PCR #1, except number of cycles is increased from 30 to 35.

9. Add 1/10 volume of sodium acetate and 2.5 volumes absolute ethanol to precipitate LM-PCR #2 product.
10. Place on ice for 30 min, then centrifuge at 4°C for 30 min at 2750 rcf. (If using 96-well plates.)
11. Using 96 well PCR plates you may remove supernatant by inverting the plate over a paper towel. Dissolve pellet in 100 µl of 3X SSC buffer or other spotting solution.

NOTE: Typical yield is 40-50 µg of DNA.

2.6.2. Protocol 2 – Post-production processing of aldehyde BAC array slides (adapted from http://www.schott.com/nexterion)

Equipment and reagents

- 10% SDS
- $NaBH_4$
- absolute ethanol
- 10X PBS
- centrifuge or blower
- water bath
- slide holder

Method

1. After spotting, let arrays dry in array printer >2 h.
2. Remove printed arrays from array printer and place into wash containers.
3. Wash arrays in 0.1% SDS solution twice for 2 min each wash.
4. Rinse arrays in dH_2O for 1 min.
5. Transfer the arrays to a 1X phosphate buffered saline solution containing 0.5% $NaBH_4$ and 20% absolute ethanol.
6. Agitate arrays for 3 min.
7. Rinse in dH_2O for 1 min.
8. Place arrays into boiling water for 30 sec.
9. Rinse at room temperature in dH_2O for 1 min.
10. Dry arrays by centrifugation in 50 ml conical tubes (no lids) for 5 min at 700 rcf, or oil-free air stream.
11. Store in a desiccated, vacuum-sealed bag in the dark until required for hybridization.

2.6.3. Protocol 3 – BAC array hybridization protocols

Equipment and reagents[7]

- Appropriate reference DNA
- Sample DNA
- 10% SDS
- 20X SSC Buffer
- sodium acetate (3.0 M, pH 5.5)
- absolute ethanol
- cyanine dyes
    - cyanine-3-dCTP (1 mM; Perkin Elmer)
    - cyanine-5-dCTP (1 mM; Perkin Elmer)
- Klenow (9 U/µl; Promega)
- Kenow buffer (10X; Promega)
- 10X "Random priming dNTP mix" (Invitrogen)
    - 2 mM dATP
    - 2 mM dGTP
    - 2 mM dTTP
    - 1.2 mM dCTP
- Random octamers (150 µg/µl stock; Alpha DNA)
- 1X DIG Easy hybridization buffer (4.617 g in 10 ml dH_2O; Roche)
- Sheared herring Sperm DNA (20 µg/µl; Invitrogen)
- yeast tRNA (10 µg/µl; Calbiochem)
- C_0t-1 DNA (1 µg/µl; Roche/Invitrogen)
- microfuge
- centrifuge or blower
- incubator at 45°C
- incubator at 37°C
- heating block at 45°C

[7] Protocol adapted from Ishkanian et al. (2004).

- heating block at 85°C
- Coplin jar or slide staining boxes
- Thermocycler
- Columns:
    - ProbeQuant Sephadex G-50 columns (Amersham)

    or

    - YM-30 Microcons (MilliQ)
- Spectrophotometer
- Hybridization cassettes (Telchem)
- Coverslips (24 x 60 mm; Fisher)
- BAC Array

Method

Probe generation and labeling

1. For separate reference and test sample reactions, combine:
    - DNA (50-400 ng).
    - 0.25 µl random octamers.[8]
    - 2.5 µl Klenow buffer. (Dilute to final reaction volume of 16.5 µl with dH_2O.)
2. Seal tube and boil for 10 min at 100°C.
3. Transfer immediately to ice for 2 min.
4. Add 3.75 µl of "10X Random priming dNTP mix"
5. Cyanine dyes:
    - Add 2 µl of cyanine-3-dCTP to either reference or test DNA sample.
    - Add 2 µl of cyanine-5-dCTP to the other DNA sample
6. Add 2.5 µl Klenow and mix gently. (Final reaction volume is 25 µl.)
7. Incubate at 37°C 18-36 h.
8. Remove unincorporated cyanine dyes using G-50 column or Microcon YM-30.
9. Combine labeling reactions.

[8] This value altered from original text, which stated 0.75 µl of random octamers should be added.

10. Assess incorporation values using spectrophotometer.

NOTE: Incorporation values below 3.0 pmol/µl (for 50 µl combined labeling reactions) in either channel have shown variable results. We typically measure incorporations in the 8-25 pmol/µl range, with cyanine-3 typically having higher values.

Preparation of probe for hybridization

11. Add 100 µl of C_0t-1 DNA.
12. Add 1/10 volume of sodium acetate and 2.5 volumes absolute ethanol.

NOTE: DNA pellet should appear as purple after centrifugation.

13. Resuspend the pellet in 45 µl of hybridization solution:
    - 36 µl 1X DIG Easy hybridization buffer
    - 4.5 µl sheared Herring Sperm DNA
    - 4.5 µl yeast tRNA
14. Denature probe at 85°C for 10 min.
15. Place probe at 45°C for 1 h (allows C_0t-1 annealing)

Hybridization

16. Place 45 µl probe solution onto the coverslip (or array) (Figure 2.4).
17. Gently lower the array (or coverslip), over the probe solution.
18. Place the slide into a hybridization cassette pre-warmed to 45°C.

NOTE: Add the appropriate amount of dH_2O to hybridization cassette to generate 100% humidity. (10 µl for Telechem hybridization cassettes)

19. Incubate for 36-40 h at 45°C.

Washing and scanning

20. Pre-warm wash solution (0.1X SSC 0.1%SDS) to 45°C.
21. Remove the coverslip (Figure 2.4) and place slides in wash solution.
22. Perform five sequential 5 min washes with agitation.

- Washes can be done in a Coplin jar or slide staining box.

21. Rinse arrays in 0.1X SSC a total of 5 times.

22. Dry the slides with an air stream (oil free) or by centrifuging the slides in 50 ml conical tubes at 700 rcf for 5 min (do not use lids).

23. Store the slides in the dark until scanning (signal intensities will diminish over time).

The authors strongly recommend that BAC array production and experiments be performed in a dedicated room where ozone, light exposure, and humidity can be carefully controlled.

2.7. Troubleshooting

2.7.1. BAC Array production (or spot evaluation of propidium iodide stained arrays)

- Target spot size is impacted by numerous factors.
  - The hydrophobicity of the slide chemistry as well as the wetting properties of the spotting solution (including viscosity due to DNA concentration) drastically affects spot size. Choosing a more hydrophobic substrate or more viscous/lower spreading spotting solution will yield smaller spots. The humidity at which spotting is performed affects the wetting properties of the substrate. Optimizing spotter humidity between 50-60% will produce the best results.
  - The impact speed as well as the amount of time the printing pins contact the substrate can change spot size, therefore reduction of pin over-stroke and setting the attack speed to manufacturer recommended settings may improve spot size.
  - The size of the printing pin tip is the most significant factor in the final spot size. Refer to your pin suppliers specifications to select the appropriate product for your application.

- Post-spotting treatment of arrays can also affect spot size, therefore careful reference to slide manufacturer protocols is needed.

- Missing target spots on the array.
  - This can results from clogging or damaged pins that may, over time, have developed a coating of material that reduces their performance. Replace or clean according to manufacturer's protocol.
  - Most printheads allows printing pins to move vertically upon impaction with the substrate, therefore users should check that pins move freely and do not stick.
  - Check that spotting solution source plates contain a volume sufficient to allow pins to contact the solution, or that the spotter is calibrated to allow pins to contact the solution.
  - Also ensure that pin washing system is functioning properly during print run as failure to properly clean pins can cause significant decrease in spot quality over time.

2.7.2. DNA sample troubleshooting

- Contaminating RNA in DNA samples will raise the absorbance 260/280 ratio (affecting concentration assessment) and potentially impact downstream reactions. Treatment of sample with RNase A enzyme will remove RNA (samples will need to be extracted again with organic solvents – phenol, chloroform).
- Contaminating proteins will lower absorbance 260/280 ratio (affecting concentration assessment) and potentially impact downstream reactions. Extracting a sample again with organic solvents will remove unwanted proteins.
- Residual organic solvents will also affect the 260/280 ratio and downstream reactions. Repeat organic solvent extraction, being careful not to disrupt aqueous:organic interface.
- High sample salt concentration can impact downstream enzyme function. Even for samples with anticipated low yield of DNA, apply *multiple* 70% EtOH washes of pellet during DNA extraction to reduce salt concentration.

- If the average fragment size in a DNA sample is too small, subsequent enzyme reactions may not be effective. If there is sufficient material, run small amount of sample against standards to determine size range. Alternatively, use low DNA yield quantification methods, such as Randomly Amplified Polymorphic DNA (RAPD)-PCR (Siwoski et al. 2002).

2.7.3. Post-labeling troubleshooting

- The amount of probe generated may be inefficient if DNA is not fully single stranded, if Klenow enzyme is inhibited by sample contaminants, or if template DNA is degraded (see "DNA sample troubleshooting", section 2.7.2). Sites of DNA damage, such as thymidine dimers and abasic sites caused by various fixatives may not be passable by the Klenow enzyme and fragmented samples may only produce short probes. To avoid this, ensure that probes are only generated from test and reference samples meeting minimum size and quality standards.
- Low dye incorporations may result from contamination of samples (with EtOH, organic solvents, proteins, etc. – see above), the presence of low quality cyanine dyes, a poor cold:hot nucleotide ratio, or poor coupling of cyanine to nucleotide. To avoid this, ensure that test and reference samples are clean and free of contaminants. Please note that cold:hot nucleotide ratios may need to be determined empirically.

2.7.4. Post-scanning troubleshooting

- Dim slides may results from numerous events.
    - Overly aggressive post-hybridization washing can remove bound probe and can be countered by reducing wash stringency (i.e. decreasing wash temperature or wash time).
    - Dimness due to short exposure times or low sensitivity settings for scanners may be avoided by altering scan settings if possible.

    - Undertaking array experiments in a controlled environment may compensate for dimness owing to environmental factors that degrade cyanine dyes such as excess light or ozone.
    - Dim images may also result from the DNA content of the spotted target being too low. This can occur because of low DNA concentration in spotting solution or removal of too much DNA during post-printing treatment. Close monitoring of these factors will resolve this issue.
- Slides can be overly bright if there is inefficient repeat blocking by C_0t-1 DNA. If C_0t-1 DNA is of low quality, or if too little Cot-1 DNA is used, repeats will not be blocked effectively, and the amount of probe binding to the array will be greatly increased. To counter this, ensure sufficient amount of C_0t-1 DNA is used.
- Slide gradients can result from numerous factors
    - If the solution on the array dries out during hybridization, a gradient can arise. To prevent this, ensure hybridization chambers have sufficient humidity and that an adequate volume of probe solution is applied to the array.
    - If performing hybridization with a coverslip, ensure that the covered array is not touching the sides of hybridization chamber as this will result in probe solution being drawn out from under the coverslip.
    - Gradients can also result during washing of arrays, causing unbound probe to be removed unevenly. This can be addressed by ensuring that wash solutions fully cover entire slide.
- Dust or precipitated probe can cause speckles in scanned images. Rinsing in 0.1X SSC or drying with compressed air may remove these.
- High scan background, caused by inefficient slide washing or incomplete slide pre-treatment, can reduce the amount of usable data obtained from a slide. Increasing wash stringency may improve this. Alternatively, if incomplete reduction of aldehyde to alcohol groups on slide surface is the problem, the reducing agent used in slide processing may need to be replaced.

2.7.5. Post-visualization troubleshooting

- Occasionally, completely unrelated samples may yield identical genomic profiles. Such profiles may be caused by several factors including test DNA samples that are degraded or of limited abundance.
- "Noisy" hybridizations (i.e. hybridizations with a wide range can have multiple causes)
    - "Noisy" hybridizations may be an artifact of a dim overall hybridization. Low overall intensity will lead to an increase in required exposure (or gain) by the scanner, a process which can introduce noise.
    - Poor DNA quality may also contribute to noise. Ensure that test and reference DNA is clean and of high molecular weight.
    - Noise in hybridizations may also be the result of a poor representational amplification being used.
- Low signal intensity ratios may be present due to poor suppression of repeats by C_0t-1 DNA. It is best to evaluate C_0t-1 using a sample with known alterations. Heterogeneity of sample (e.g. normal cell contamination in isolated tumor cells) may also cause a reduction in apparent ratios. To avoid this, ensure that rigorous microdissection is undertaken.

2.8. Accrual of clinical lung tumour specimens for molecular analysis

Multiple challenges can affect molecular analysis of clinical tumour specimens, including proper collection and storage of tissues (whether archiving fixed biopsied specimens or freezing freshly resected specimens), minimization of the impact non-tumour tissue has on analysis, and extraction of DNA and RNA of sufficient quality and quantity for experimentation. These issues were addressed by carefully optimizing collection approaches for a variety of tissue sources. Subsequent chapters were based on application of these approaches.

2.8.1. Collection of fresh surgically resected tumour specimens

Nearly all previous analyses of lung tumour genome alterations have been undertaken on resected specimens that have not been fixed with formalin (Choi et al. 2007; Dehan

et al. 2007; Gallegos Ruiz et al. 2008; Kang et al. 2008; LaFramboise et al. 2005; Lo et al. 2008; Nymark et al. 2006; Peng et al. 2005; Shibata et al. 2005; Tonon et al. 2005; Weir et al. 2007; Zhao et al. 2005). In our hands, portions of these tumours were removed and placed onto dry ice immediately following surgery to limit degradation of nucleic acids. These samples were fit into cryomolds and embedded in Optimal Cutting Temperature compound (Tissue-Tek) for storage in a liquid N_2 freezer (approximately -180°C). Downstream processing for each case involved cutting serial 5 µm sections on a cryotome, the first and last sections cut placed on slides and stained with hematoxylin and eosin (H&E) to visualize tumour-dense regions. Approximately half of the intermediate sections that were cut were reserved for DNA extraction, with the other half for RNA extraction. Cross-sections used for DNA extraction underwent ethanol fixation and were stored at -80°C until microdissected. Cross-sections saved for RNA extraction were first treated with RNAlater and kept at 4°C for ~3 hours, then transferred to -80°C as well. These careful practices generally yielded high quality DNA and RNA following careful microdissection (Figure 2.6).

With respect to microdissection, targeted regions typically had ≥70% tumour cells. Although earlier reports suggested that the BAC array CGH platform in use was able to detect alterations in DNA with as much as 50% non-tumour cell content, in practical application we found that less stringent microdissection practices could lead to a marked decrease in the ability to detect genomic alterations (Figure 2.7) (Garnis & Coe et al. 2005b). Where regions identified on H&E slides during pathology review were sufficiently large, manual approaches were used. Typically, a 20G needle was used to scrape and collect cells within these areas. Given ~6 pg of DNA per cell, NSCLC tumour cells averaging ~20-50 µm in diameter, and captured areas that were quite often >1 cm^2, it is unsurprising that several micrograms of DNA were often obtained (Sambrook et al. 1989; Travis 2002). For cases where tissue heterogeneity or targeted region size precluded microdissection, laser-capture microdissection was used (Cellcut System, Molecular Machines & Industries) (Figure 2.8).

2.8.2. Collection of fixed lung tumour specimens from sample archives

Patients with late stage lung tumours are not typically treated by surgery; fixed biopsy specimens collected by medistinoscopy represent the only tissue available for research purposes. As mentioned above, some high throughput genome profiling platforms are unable to provide usable data for DNA isolated from tumour archives. This can result because of insufficient DNA quantity, particularly where small early stage lesions are being analyzed, however it is more typically due to the poor quality of DNA isolated from archival samples (irreversible damage may be caused by formalin fixation). Consequently, only a limited number of formalin-fixed paraffin-embedded (FFPE) lung cancer cases have undergone whole genome profiling (Aviel-Ronen et al. 2008; Gallegos Ruiz et al. 2007; Buys et al. 2009). An optimized DNA extraction protocol based on standard phenol:chloroform extraction was used to maximize DNA quality and yield. Although DNA from FFPE samples was typically more fragmented than DNA isolated from frozen tissues (see Figure 2.6), good genomic profiles could generally be expected where the average fragment size was >500 bp. While spurious alterations may be detected in DNA from FFPE specimens (Mc Sherry et al. 2007), a direct comparison of frozen and FFPE tissues produced nearly identical genomic alteration results (Figure 2.9).

Figure 2.1

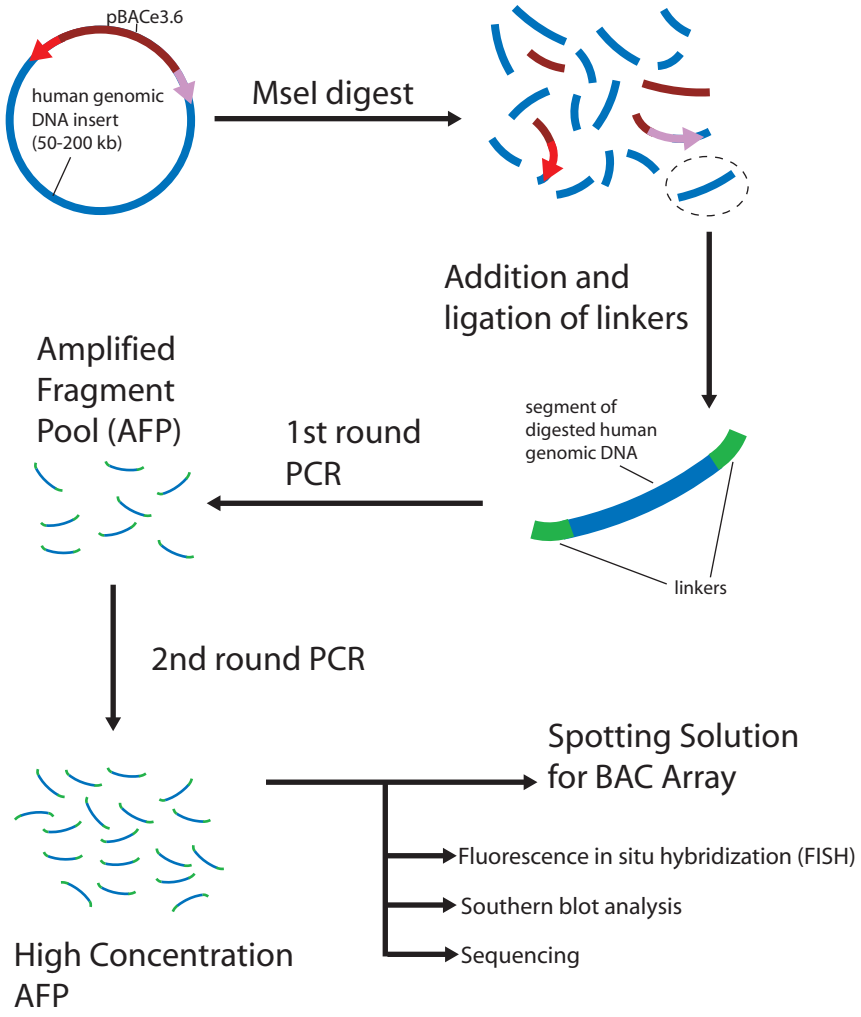


Figure 2.1 – Linker-mediated PCR (LM-PCR) mediated production of amplified fragment pools (AFP) of BAC clone DNA for use in array synthesis. BAC clone DNA is isolated and digested with MseI to release the human genomic insert from the vector. Linkers are then ligated to the MseI-digested ends of the genomic insert and amplified in a linker-mediated PCR reaction (LM-PCR). The remaining product is further amplified by another round of LM-PCR which is used to verify the clone identity and the creation of spotting solution for array production.

Figure 2.2

— test DNA source

Single male reference	- This reference should have polymorphisms characterized in advance of experiments - Could allow characterization of alterations on sex chromosomes Potential problems: absence of gender mismatch limits the use of the X chromosome to determine relative separation of defined DNA copy number differences
Pooled male reference	- Polymorphisms (somewhat) diluted by pooled source - Readily available commercially Potential problems: changes in individuals contributing to sample pool contributing population not sufficiently large to dilute polymorphisms absence of gender mismatch limits the use of the X chromosome to determine relative separation of defined DNA copy number differences
Single female reference	- This reference should have polymorphisms characterized in advance of experiments - Gender mismatch allows determination of single copy shift between X chromosomes (serves as a control) Potential problems: no data for Y chromosome clones
Pooled female reference	- Polymorphisms (somewhat) diluted by pooled source - Readily available commercially - Gender mismatch allows determination of single copy shift between X chromosomes (serves as a control) Potential problems: changes in contributors contributing to sample pool contributing population not sufficiently large to dilute polymorphisms no data for Y chromosome clones
Matched normal	- Reduces experimental noise resulting from slight sequence variation (unpublished results) - Excellent reference if test sample is derived from discrete abnormal portion of test subject (e.g. a tumor) Potential problems: self vs self hybridizations not usable for assessment of consitutional DNA differences

Figure 2.2 – Types of reference DNA sample for competitive hybridization experiments. Genomic results for lung tumors described elsewhere in this thesis were performed against the same individual male reference DNA sample. This sample was chosen because its copy number polymorphisms had already been characterized and the anonymous volunteer could be approached again if more DNA was needed.

Figure 2.3 – Genome labeling approaches. A. Genome sampling vs. whole genome labeling. Labeled DNA probes are shown as grey circles demonstrating the differential genome coverage of both methods. B. Random prime method of whole genome labeling for use in BAC array CGH experiments. (i) Genomic DNA from both test and reference samples are individually added to a mixture containing random primers and buffer solution and heated to 100°C in order denature the double strands. (ii) The solution is then snap cooled on ice in order to keep the DNA single stranded. dNTPs, Klenow enzyme, and either cyanine-3 or -5 dCTPs are added to the DNA and the resulting mixture is placed at 37°C. (iii) The random primers then anneal to the single stranded DNA template leaving a 3'-OH that allows Klenow to start DNA synthesis by incorporating both dNTPs and cyanine-labeled dCTPs. (iv) Klenow continues moving along the DNA template displacing previously synthesized stands resulting in linear amplification of the starting material. (v) The resulting probe contains incorporated cyanine dyes allowing its use in array CGH experiments. C. Differences between direct and indirect labeling. (i) With direct labeling, cyanine-coupled nucleotides are directly incorporated into generated probe. (ii) When performing indirect labeling, amino-allyl-coupled nucleotides are incorporated into the probe. Fluorophores are then bound to the nucleotides in a second step.

Figure 2.3

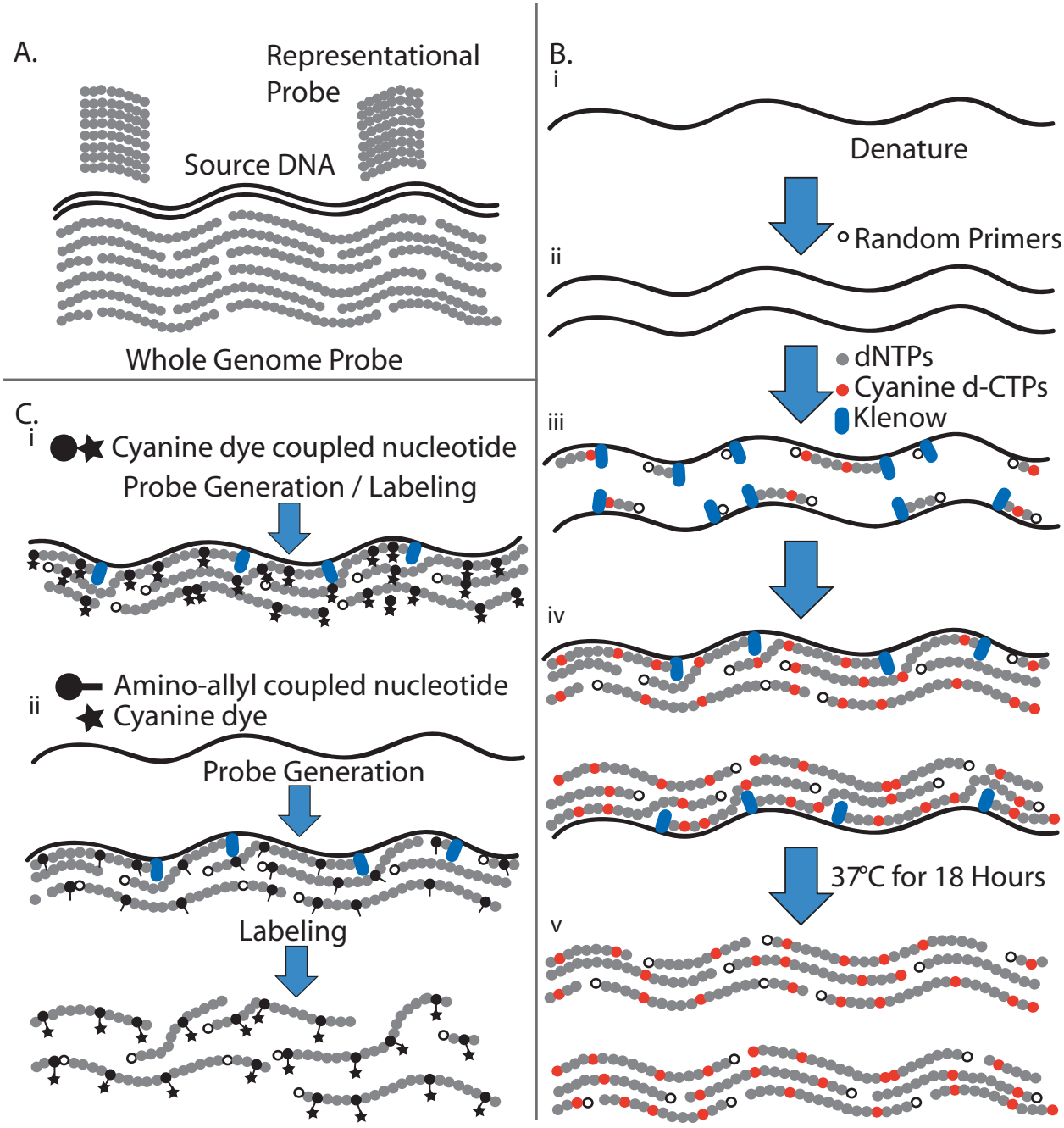

A.
Representational Probe
Source DNA
Whole Genome Probe
B.
i
Denature
Random Primers
ii
dNTPs
Cyanine d-CTPs
Klenow
iii
iv
37°C for 18 Hours
v
C.
i
Cyanine dye coupled nucleotide
Probe Generation / Labeling
ii
Amino-allyl coupled nucleotide
Cyanine dye
Probe Generation
Labeling

Figure 2.4 – Coverslip application and removal for slide washing. *Coverslip Application*. A. Deposit hybridization solution onto the array surface. B. Orient the coverslip with the array. C. Place the coverslip directly onto the array taking care to cover all spotted targets (marking the back of the slide with a glass etcher or pen may help with alignment). D. The hybridization solution should spread to cover the entire array (if the hybridization solution does not reach the edges of the coverslip more solution is required). Care must be taken not to introduce bubbles between the coverslip and the array as target spots within bubbles will not hybridize probes. After application of the coverslip, the array may be placed into a hybridization chamber. *Coverslip Removal and Slide Washing*. E. Place the hybridized array into wash solution for approximately 30 sec. F. Remove the array from the wash solution, and slide the coverslip so that at least 5 mm of the coverslip is off the array. The coverslip should easily slide off the slide – if it does not, soak the slide for additional time. G. Use the exposed edge to lift the coverslip off the slide. H. Place the slide back into the wash solution and start wash protocol. For both coverslip application and removal, users must ensure that they act swiftly as failure to do so could lead to hybridization solution drying out.

Figure 2.4

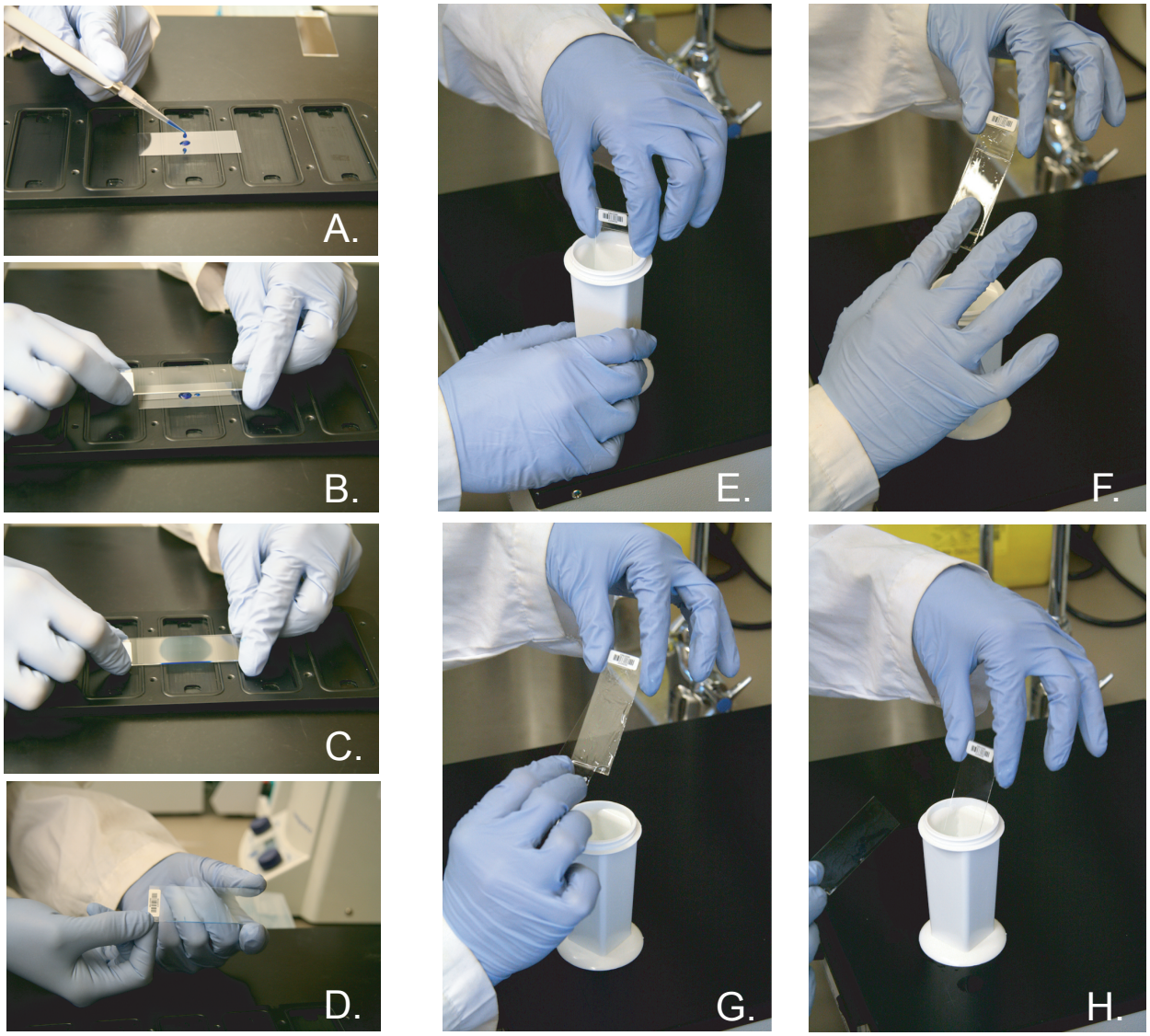


Figure 2.5

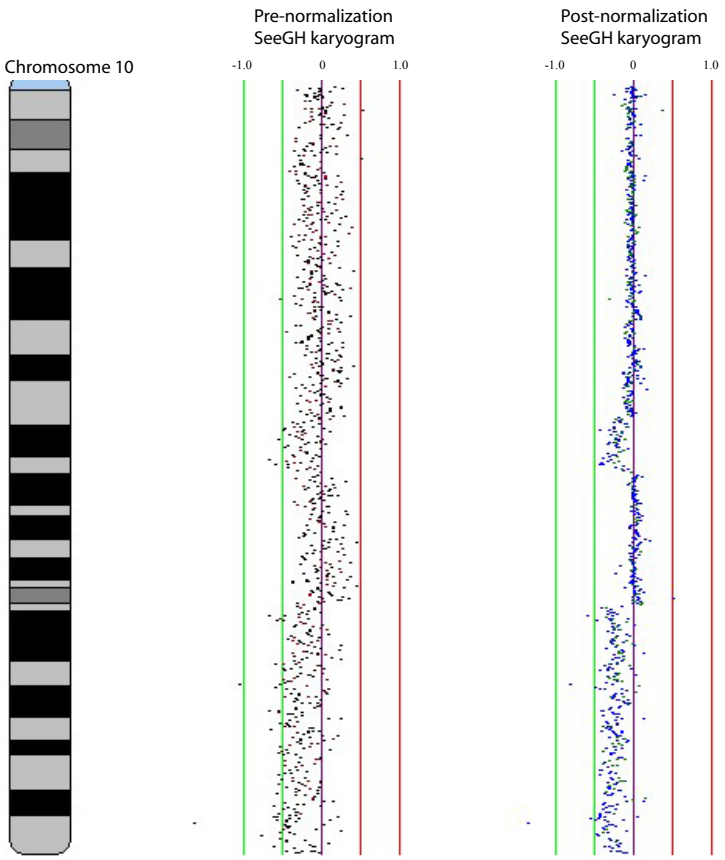


Figure 2.5 – Normalization of BAC array data. The effects of normalization of array CGH data for a competitive hybridization of a cancer cell line against its drug resistant derivative. Normalized and non-normalized $\log_2$ signal intensity ratios were plotted using SeeGH software. Clones with standard deviations among the triplicate spots >0.09 or a signal-to-noise ratio >3 were filtered from analysis. Chromosome arm 10q is represented to the left. Vertical lines denote $\log_2$ signal ratios from -1 to 1 with copy number increases to the right (red lines) and decreases to the left (green lines) of zero (purple line). A $\log_2$ signal ratio of zero represents equivalent copy number between the hybridized samples. Each black dot represents a single BAC clone. Normalization was performed using a custom normalization program with the parameters set as default (Khojasteh et al. 2005).

Table 2.1 – Useful web links.

BAC Resources		
Description		**URL**
CHORI Website (Roswell Park Clone Library Information)		http://www.chori.org/bacpac/
Caltech Genome Research Laboratory (CTD Library Information)		http://informa.bio.caltech.edu
Other Resources		
Description		**URL**
Finger Print Contig (FPC) Database		http://genome.wustl.edu/projects/human/index.php?fpc=1
UCSC Genome Mapping		http://genome.ucsc.edu/
Ensembl		http://www.ensembl.org/index.html
Academic BAC Array Providers		
Description		**URL**
Albertson Lab		http://cc.ucsf.edu/albertson/
British Columbia Cancer Research Centre BAC Array Group		http://www.bccrc.ca/cg/ArrayCGH_Group.html, www.arraycgh.ca
Fred Hutchison Cancer Research Centre		http://www.fhcrc.org
Commercial Sources		
Description	**Items available**	**URL**
Agilent	Reagents, arrays, scanners	http://www.agilent.com
Amersham	Reagents, arrays	http://www.amersham.com
API	Scanners	http://www.api.com
Corning	Slides	http://www.corning.com
Genetix	Printers, plates, reagents	http://www.genetix.com
Invitrogen	Reagents	http://www.invitrogen.com
Perkin Elmer	Spotters, scanners, reagents	http://www.perkinelmer.com
Roche	Reagents	http://www.roche-applied-science.com
Schott US	Slides	http://www.us.schott.com/english/index.html
Telechem	Pins	http://www.arrayit.com
University Health Network	Arrays	http://www.microarray.ca
BioRad	Scanners, printers	http://www.biorad.com
Vysis	Reagents, arrays, scanners	http://www.vysis.com
Spectral Genomics	Reagents, arrays, scanners	http://www.spectralgenomics.com/
Genomic Solutions	Scanners, spotters, reagents	http://www.genomicsolutions.com
Signature Genomics	Arrays	http://www.signaturegenomics.com

Figure 2.6 – Evaluating DNA and RNA quality. A. Agarose gel results for five different lung tumour DNA samples isolated from frozen resected tissues (1 kb+ standard reference ladder was run in the lane on the left). All cases exhibited high molecular weight DNA that is characteristically obtained from such tissue, the average fragment size running larger than the highest band in the reference ladder (12216 base pairs). B. Agarose gel results for five different lung tumour DNA samples isolated from FFPE tissue (1 kb+ standard reference ladder to the left). Variable fragment sizes were observed for these samples, with the degree of degradation for the DNA sample run in the fourth lane indicating strongly that subsequent labeling reactions are unlikely to succeed. C. Agarose gel results for 7 different lung tumour RNA samples obtained from frozen tissue (1 kb+ standard reference ladder to the left). Lanes 2-5 (from the left) show results for RNA samples stored at -80°C. The presence of characteristic ~1900 bp and ~5000 bp bands for 18s and 28s ribosomal RNA are indicative of good sample quality (Gonzalez et al. 1985; Gonzalez & Schmickel 1986). Lanes 6-8 show results for RNA samples stored at room temperature, with the only detectable fragments at the bottom of the gel, indicating RNA degradation.

Figure 2.6

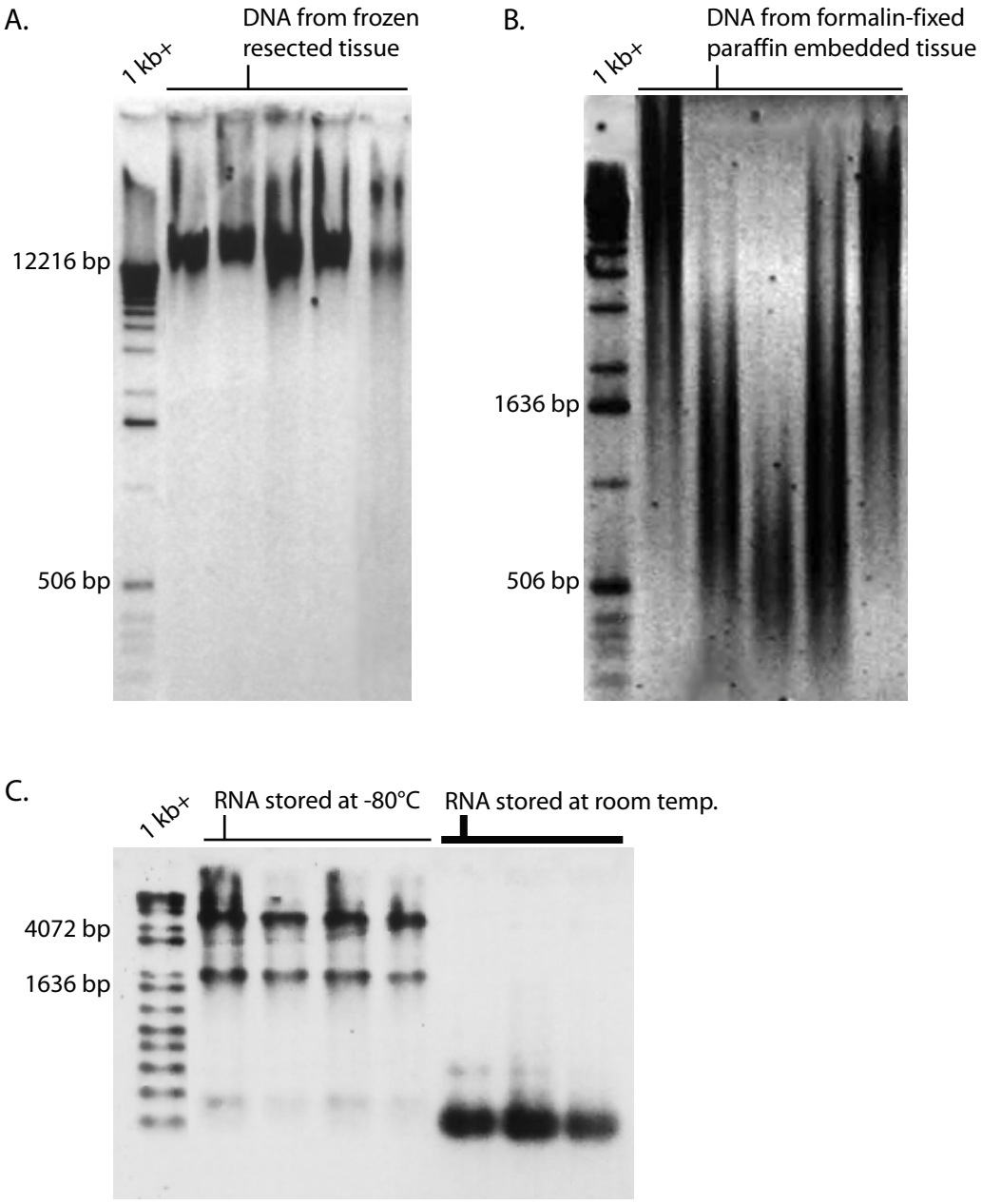

A.
DNA from frozen resected tissue
1 kb+
12216 bp
506 bp
B.
DNA from formalin-fixed paraffin embedded tissue
1 kb+
1636 bp
506 bp
C.
1 kb+
RNA stored at -80°C
RNA stored at room temp.
4072 bp
1636 bp

Figure 2.7 – Impact of tissue heterogeneity on tumour genome profiling for a case of squamous cell lung carcinoma. Successive cross-sections from the resected tumour specimen were cut, with some undergoing careful microdissection to remove adjacent normal cells and stroma prior to DNA extraction and others being collected *in toto* for DNA isolation. Genome alteration profiles were generated for each DNA extraction; a portion of chromosome 3q is shown. Results for DNA from non-microdissected tissues are represented on the left, while results for microdissected tissues are on the right. Each BAC clone on the array CGH platform is represented as a blue spot at its known chromosomal position. Clones are plotted along the horizontal axis based on a $\log_2$ transformation of the signal intensity ratios from the experiment. Vertical lines show $\log_2$ signal intensity ratios from 1 (red line) to -1 (green line), with copy number increases to the right and decreases to the left of zero (purple line). An amplification event within 3q26.33 is highlighted in orange and is apparently muted for the DNA sample with contributions from non-tumour cells. A broad low-level segmental gain spanning this entire portion of chromosome 3q is detected only for the microdissected case.

Figure 2.7

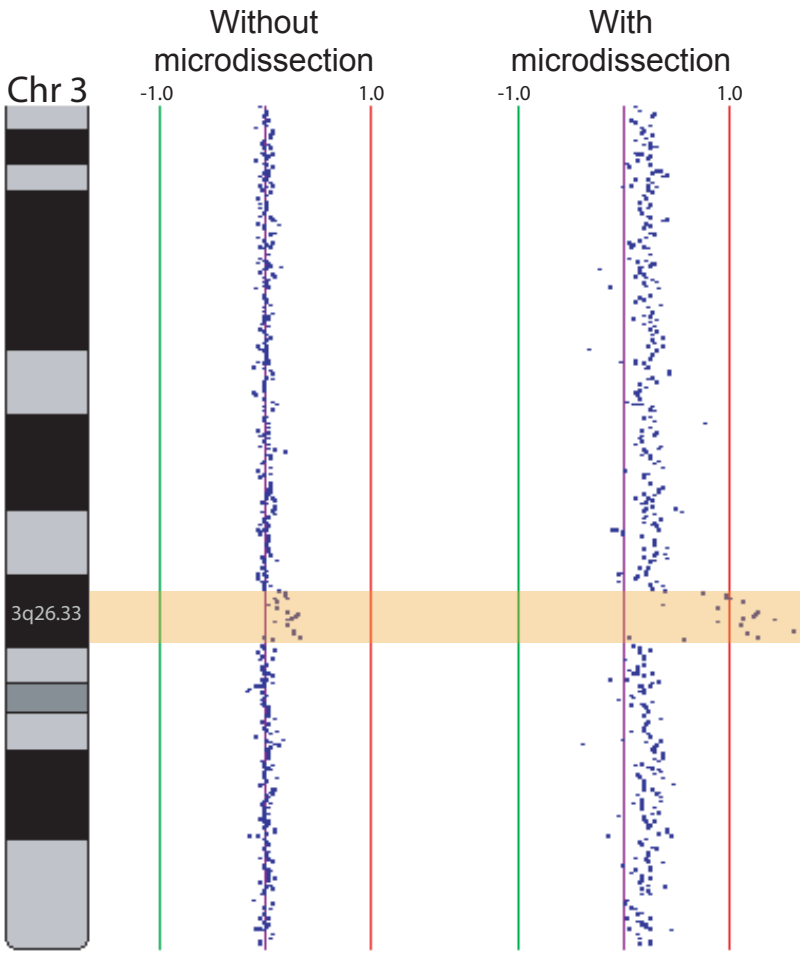

Without microdissection
With microdissection
Chr 3
-1.0
1.0
-1.0
1.0
3q26.33

Figure 2.8

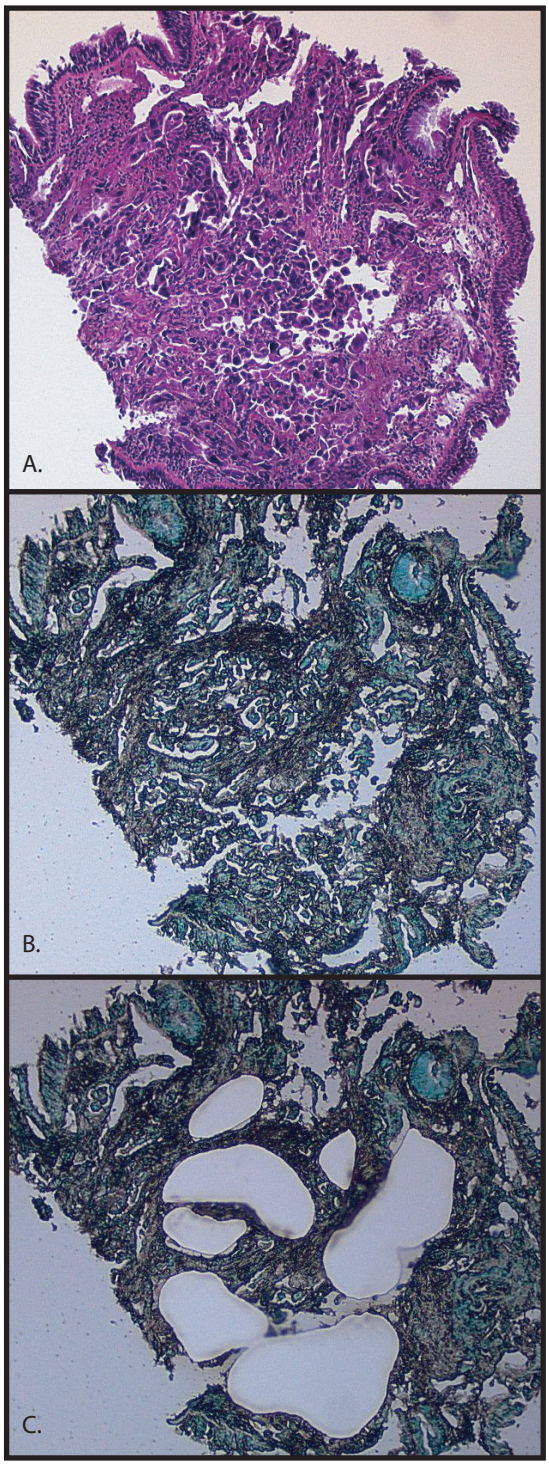


Figure 2.8 – Microdissection of for a clinical lung tumour specimen. Serial 5 µm sections were cut (all images 100X magnification). A. Image of an H&E stained reference slide to guide in identification of regions with greater tumour cell density. B. Image of subsequent tumour cross-section, pre-microdissection. Each slide cut for microdissection was stained by methyl green to improve visualization for targeted regions. C. Image of same cross-section as in B. following laser capture microdissection. Void areas were tumour-rich regions captured for DNA extraction.

Figure 2.9

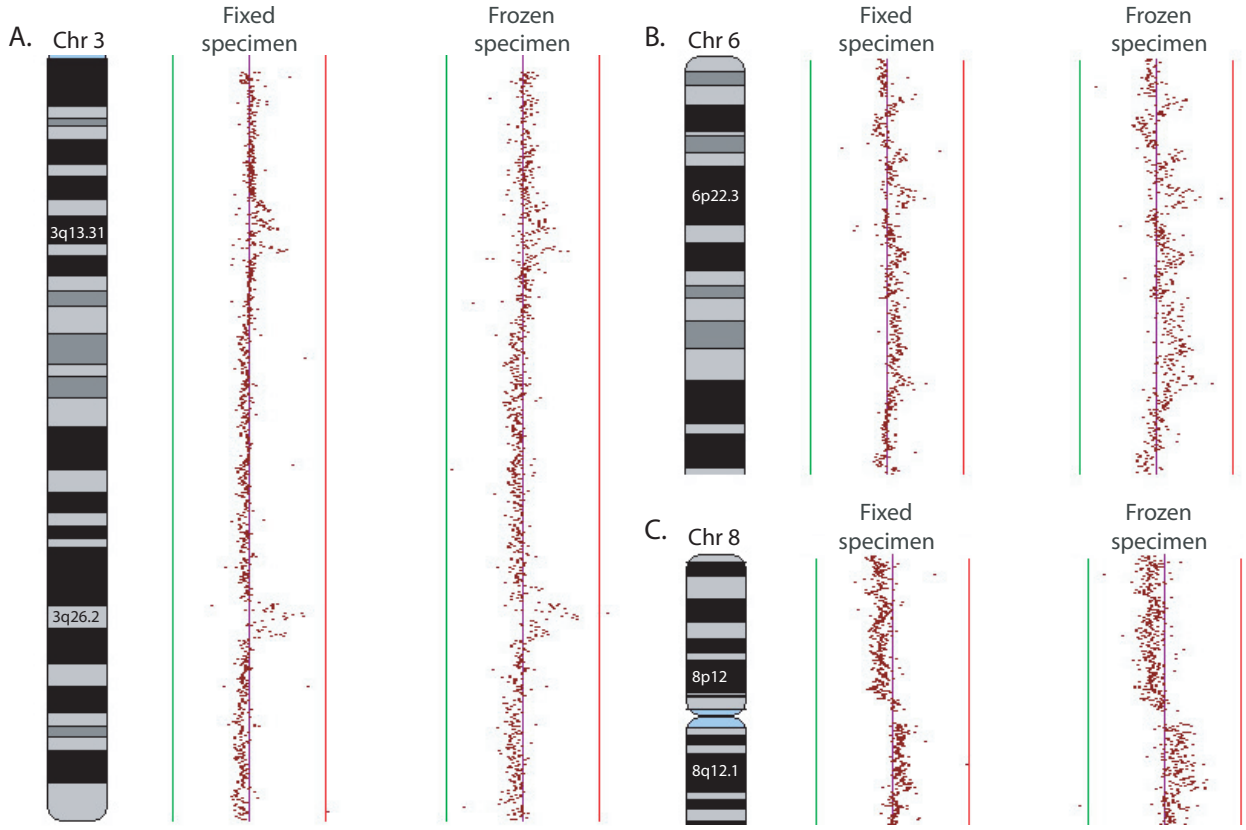


Figure 2.9 – Impact of fixation on genome profiling results. Portions of a single lung adenocarcinoma tumour were handled and stored differently; one fragment was frozen without fixation while the other was fixed in formalin, embedded in paraffin, and stored at room temperature. Identical genome alteration data were obtained regardless of the approach to storage, including complicated genomic amplifications around chromosome bands 3q13.31 and 3q26.2 (A), complex rearrangements all along chromosome arm 6p (B), and broad segmental alterations spanning all of chromosome arm 8p and a portion of 8q (C).

2.9. References

Adey, N. B., M. Lei, M. T. Howard, J. D. Jensen, D. A. Mayo, D. L. Butel, S. C. Coffin, T. C. Moyer, D. E. Slade, M. K. Spute, A. M. Hancock, G. T. Eisenhoffer, B. K. Dalley & M. R. McNeely, 2002. Gains in sensitivity with a device that mixes microarray hybridization solution in a 25-microm-thick chamber. *Anal Chem,* 74(24), 6413-7.

Amasino, R. M., 1986. Acceleration of nucleic acid hybridization rate by polyethylene glycol. *Anal Biochem,* 152(2), 304-7.

Aviel-Ronen, S., B. P. Coe, S. K. Lau, G. da Cunha Santos, C. Q. Zhu, D. Strumpf, I. Jurisica, W. L. Lam & M. S. Tsao, 2008. Genomic markers for malignant progression in pulmonary adenocarcinoma with bronchioloalveolar features. *Proc Natl Acad Sci U S A,* 105(29), 10155-60.

Bignell, G. R., J. Huang, J. Greshock, S. Watt, A. Butler, S. West, M. Grigorova, K. W. Jones, W. Wei, M. R. Stratton, P. A. Futreal, B. Weber, M. H. Shapero & R. Wooster, 2004. High-resolution analysis of DNA copy number using oligonucleotide microarrays. *Genome Res,* 14(2), 287-95.

Bonner, R. F., M. Emmert-Buck, K. Cole, T. Pohida, R. Chuaqui, S. Goldstein & L. A. Liotta, 1997. Laser capture microdissection: molecular analysis of tissue. *Science,* 278(5342), 1481,3.

Borden, J. R., C. J. Paredes & E. Papoutsakis, 2005. Diffusion, Mixing, and Associated Dye Effects in DNA-Microarray Hybridizations. *Biophys J*.

Britten, R. J. & D. E. Kohne, 1968. Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. *Science,* 161(841), 529-40.

Buckley, P. G., K. K. Mantripragada, M. Benetkiewicz, I. Tapia-Paez, T. Diaz De Stahl, M. Rosenquist, H. Ali, C. Jarbo, C. De Bustos, C. Hirvela, B. Sinder Wilen, I. Fransson, C. Thyr, B. I. Johnsson, C. E. Bruder, U. Menzel, M. Hergersberg, N. Mandahl, E. Blennow, A. Wedell, D. M. Beare, J. E. Collins, I. Dunham, D. Albertson, D. Pinkel, B. C. Bastian, A. F. Faruqi, R. S. Lasken, K. Ichimura, V. P. Collins & J. P. Dumanski, 2002. A full-coverage, high-resolution human chromosome 22 genomic microarray for clinical and research applications. *Hum Mol Genet,* 11(25), 3221-9.

Burgess, J. K., 2001. Gene expression studies using microarrays. *Clin Exp Pharmacol Physiol,* 28(4), 321-8.

Buys, T. P., S. Aviel-Ronen, T. K. Waddell, W. L. Lam & M. S. Tsao, 2009. Defining genomic alteration boundaries for a combined small cell and non-small cell lung carcinoma. *J Thorac Oncol,* 4(2), 227-39.

Carter, N. P., H. Fiegler & J. Piper, 2002. Comparative analysis of comparative genomic hybridization microarray technologies: report of a workshop sponsored by the Wellcome Trust. *Cytometry,* 49(2), 43-8.

Casey, J. & N. Davidson, 1977. Rates of formation and thermal stabilities of RNA:DNA and DNA:DNA duplexes at high concentrations of formamide. *Nucleic Acids Res,* 4(5), 1539-52.

Cheung, S. W., C. A. Shaw, W. Yu, J. Li, Z. Ou, A. Patel, S. A. Yatsenko, M. L. Cooper, P. Furman, P. Stankiewicz, J. R. Lupski, A. C. Chinault & A. L. Beaudet, 2005. Development and validation of a CGH microarray for clinical cytogenetic diagnosis. *Genet Med,* 7(6), 422-32.

Chi, B., R. J. DeLeeuw, B. P. Coe, C. MacAulay & W. L. Lam, 2004. SeeGH - A software tool for visualization of whole genome array comparative genomic hybridization data. *BMC Bioinformatics,* 5(1), 13.

Choi, Y. W., J. S. Choi, L. T. Zheng, Y. J. Lim, H. K. Yoon, Y. H. Kim, Y. P. Wang & Y. Lim, 2007. Comparative genomic hybridization array analysis and real time PCR reveals genomic alterations in squamous cell carcinomas of the lung. *Lung Cancer,* 55(1), 43-51.

Cleveland, W. S., 1979. Robust locally weighted regression and smoothing scatterplots. *J. Amer. Stat. Assoc.,* 74, 829-36.

Coe, B. P., L. J. Henderson, C. Garnis, M. S. Tsao, A. F. Gazdar, J. Minna, S. Lam, C. Macaulay & W. L. Lam, 2005. High-resolution chromosome arm 5p array CGH analysis of small cell lung carcinoma cell lines. *Genes Chromosomes Cancer,* 42(3), 308-13.

Conrad, W., (2005). 100,000 SNP Copy Number Studies Find New Genes Linked to Cancer, in *Affymetrix Microarray Bulletin* Affymetrix, Inc., 1-24.

Davies, J. J., I. M. Wilson & W. L. Lam, 2005. Array CGH technologies and their applications to cancer genomes. *Chromosome Res,* 13(3), 237-48.

Davison, E. J., P. S. Tarpey, H. Fiegler, I. P. Tomlinson & N. P. Carter, 2005. Deletion at chromosome band 20p12.1 in colorectal cancer revealed by high resolution array comparative genomic hybridization. *Genes Chromosomes Cancer*.

de Vries, B. B., R. Pfundt, M. Leisink, D. A. Koolen, L. E. Vissers, I. M. Janssen, S. Reijmersdal, W. M. Nillesen, E. H. Huys, N. Leeuw, D. Smeets, E. A. Sistermans, T. Feuth, C. M. van Ravenswaaij-Arts, A. G. van Kessel, E. F. Schoenmakers, H. G. Brunner & J. A. Veltman, 2005. Diagnostic genome profiling in mental retardation. *Am J Hum Genet,* 77(4), 606-16.

Dehan, E., A. Ben-Dor, W. Liao, D. Lipson, H. Frimer, S. Rienstein, D. Simansky, M. Krupsky, P. Yaron, E. Friedman, G. Rechavi, M. Perlman, A. Aviram-Goldring, S. Izraeli,

M. Bittner, Z. Yakhini & N. Kaminski, 2007. Chromosomal aberrations and gene expression profiles in non-small cell lung cancer. *Lung Cancer,* 56(2), 175-84.

Emmert-Buck, M. R., R. F. Bonner, P. D. Smith, R. F. Chuaqui, Z. Zhuang, S. R. Goldstein, R. A. Weiss & L. A. Liotta, 1996. Laser capture microdissection. *Science,* 274(5289), 998-1001.

Fare, T. L., E. M. Coffey, H. Dai, Y. D. He, D. A. Kessler, K. A. Kilian, J. E. Koch, E. LeProust, M. J. Marton, M. R. Meyer, R. B. Stoughton, G. Y. Tokiwa & Y. Wang, 2003. Effects of atmospheric ozone on microarray data quality. *Anal Chem,* 75(17), 4672-5.

Feinberg, A. P. & B. Vogelstein, 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.,* 132, 6-13.

Feinberg, A. P. & B. Vogelstein, 1984. "A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity". Addendum. *Anal Biochem,* 137(1), 266-7.

Fiegler, H., P. Carr, E. J. Douglas, D. C. Burford, S. Hunt, C. E. Scott, J. Smith, D. Vetrie, P. Gorman, I. P. Tomlinson & N. P. Carter, 2003. DNA microarrays for comparative genomic hybridization based on DOP-PCR amplification of BAC and PAC clones. *Genes Chromosomes Cancer,* 36(4), 361-74.

Frengen, E., D. Weichenhan, B. Zhao, K. Osoegawa, M. van Geel & P. J. de Jong, 1999. A modular, positive selection bacterial artificial chromosome vector with multiple cloning sites. *Genomics,* 58(3), 250-3.

Gallegos Ruiz, M. I., K. Floor, P. Roepman, J. A. Rodriguez, G. A. Meijer, W. J. Mooi, E. Jassem, J. Niklinski, T. Muley, N. van Zandwijk, E. F. Smit, K. Beebe, L. Neckers, B. Ylstra & G. Giaccone, 2008. Integration of gene dosage and gene expression in non-small cell lung cancer, identification of HSP90 as potential target. *PLoS ONE,* 3(3), e0001722.

Gallegos Ruiz, M. I., H. van Cruijsen, E. F. Smit, K. Grunberg, G. A. Meijer, J. A. Rodriguez, B. Ylstra & G. Giaccone, 2007. Genetic heterogeneity in patients with multiple neoplastic lung lesions: a report of three cases. *J Thorac Oncol,* 2(1), 12-21.

Garnis, C., C. Baldwin, L. Zhang, M. P. Rosin & W. L. Lam, 2003. Use of complete coverage array comparative genomic hybridization to define copy number alterations on chromosome 3p in oral squamous cell carcinomas. *Cancer Res,* 63(24), 8582-5.

Garnis, C., T. P. H. Buys & W. L. Lam, 2004a. Genetic alteration and gene expression modulation during cancer progression. *Mol Cancer,* 3(1), 9.

Garnis, C., J. Campbell, J. J. Davies, C. Macaulay, S. Lam & W. L. Lam, 2005a. Involvement of multiple developmental genes on chromosome 1p in lung tumorigenesis. *Hum Mol Genet,* 14(4), 475-82.

Garnis, C., J. Campbell, L. Zhang, M. P. Rosin & W. L. Lam, 2004b. OCGR array: an oral cancer genomic regional array for comparative genomic hybridization analysis. *Oral Oncol,* 40(5), 511-9.

Garnis, C., B. Coe, L. J. Henderson, A. Ishkanian, S. Watson, M. Marra, J. Minna, S. Lam, C. MacAulay & W. Lam, 2004c. Construction and optimization of chromosome arm-specific comparative genomic hybridization arrays for identifying genetic alterations in preinvasive lung cancers. *Chest,* 125(5 Suppl), 104S-5S.

Garnis, C., B. P. Coe, S. L. Lam, C. MacAulay & W. L. Lam, 2005b. High-resolution array CGH increases heterogeneity tolerance in the analysis of clinical samples. *Genomics,* 85(6), 790-3.

Garnis, C., J. J. Davies, T. P. Buys, M. S. Tsao, C. MacAulay, S. Lam & W. L. Lam, 2005c. Chromosome 5p aberrations are early events in lung cancer: implication of glial cell line-derived neurotrophic factor in disease progression. *Oncogene,* 24(30), 4806-12.

Garnis, C., C. MacAulay, S. Lam & W. Lam, 2004d. Genetic alteration on 8q distinct from MYC in bronchial carcinoma in situ lesions. *Lung Cancer,* 44(3), 403-4.

Gonzalez, I. L., J. L. Gorski, T. J. Campen, D. J. Dorney, J. M. Erickson, J. E. Sylvester & R. D. Schmickel, 1985. Variation among human 28S ribosomal RNA genes. *Proc Natl Acad Sci U S A,* 82(22), 7666-70.

Gonzalez, I. L. & R. D. Schmickel, 1986. The human 18S ribosomal RNA gene: evolution and stability. *Am J Hum Genet,* 38(4), 419-27.

Greshock, J., T. L. Naylor & A. Margolin, 2004. 1-Mb resolution array-based comparative genomic hybridization using a BAC clone set optimized for cancer gene analysis. *Genome Res,* 14(1), 179-87.

Hanahan, D. & R. A. Weinberg, 2000. The hallmarks of cancer. *Cell,* 100(1), 57-70.

Hegde, P., R. Qi, K. Abernathy, C. Gay, S. Dharap, R. Gaspard, J. E. Hughes, E. Snesrud, N. Lee & J. Quackenbush, 2000. A concise guide to cDNA microarray analysis. *Biotechniques,* 29(3), 548-50, 52-4, 56 passim.

Henderson, L. J., B. P. Coe, E. H. Lee, L. Girard, A. F. Gazdar, J. D. Minna, S. Lam, C. MacAulay & W. L. Lam, 2005. Genomic and gene expression profiling of minute alterations of chromosome arm 1p in small-cell lung carcinoma cells. *Br J Cancer,* 92(8), 1553-60.

Hughes, S., N. Arneson, S. Done & J. Squire, 2005. The use of whole genome amplification in the study of human disease. *Prog Biophys Mol Biol,* 88(1), 173-89.

Hyman, E., P. Kauraniemi, S. Hautaniemi, M. Wolf, S. Mousses, E. Rozenblum, M. Ringner, G. Sauter, O. Monni, A. Elkahloun, O. P. Kallioniemi & A. Kallioniemi, 2002. Impact of DNA amplification on gene expression patterns in breast cancer. *Cancer Res,* 62(21), 6240-5.

Iafrate, A., L. Feuk & M. Rivera, 2004. Detection of large-scale variation in the human genome. *Nat Genet,* 36(9), 949-51.

Ishkanian, A. S., C. A. Malloff, S. K. Watson, R. J. DeLeeuw, B. Chi, B. P. Coe, A. Snijders, D. G. Albertson, D. Pinkel, M. A. Marra, V. Ling, C. MacAulay & W. L. Lam, 2004. A tiling resolution DNA microarray with complete coverage of the human genome. *Nat Genet,* 36(3), 299-303.

Kameoka, Y., H. Tagawa, S. Tsuzuki, S. Karnan, A. Ota, M. Suguro, R. Suzuki, M. Yamaguchi, Y. Morishima, S. Nakamura & M. Seto, 2004. Contig array CGH at 3p14.2 points to the FRA3B/FHIT common fragile region as the target gene in diffuse large B-cell lymphoma. *Oncogene,* 23(56), 9148-54.

Kang, J. U., S. H. Koo, K. C. Kwon, J. W. Park & J. M. Kim, 2008. Gain at chromosomal region 5p15.33, containing TERT, is the most frequent genetic event in early stages of non-small cell lung cancer. *Cancer Genet Cytogenet,* 182(1), 1-11.

Khojasteh, M., W. L. Lam, R. K. Ward & C. MacAulay, 2005. A Stepwise Framework for the Normalization of Array CGH Data. *BMC Bioinformatics,* submitted.

Kohlhammer, H., C. Schwaenen, S. Wessendorf, K. Holzmann, H. A. Kestler, D. Kienle, T. F. Barth, P. Moller, G. Ott, J. Kalla, B. Radlwimmer, A. Pscherer, S. Stilgenbauer, H. Dohner, P. Lichter & M. Bentz, 2004. Genomic DNA-chip hybridization in t(11;14)-positive mantle cell lymphomas shows a high frequency of aberrations and allows a refined characterization of consensus regions. *Blood,* 104(3), 795-801.

Krzywinski, M., I. Bosdet, D. Smailus, R. Chiu, C. Mathewson, N. Wye, S. Barber, M. Brown-John, S. Chan, S. Chand, A. Cloutier, N. Girn, D. Lee, A. Masson, M. Mayo, T. Olson, P. Pandoh, A. L. Prabhu, E. Schoenmakers, M. Tsai, D. Albertson, W. Lam, C. O. Choy, K. Osoegawa, S. Zhao, P. J. de Jong, J. Schein, S. Jones & M. A. Marra, 2004. A set of BAC clones spanning the human genome. *Nucleic Acids Res,* 32(12), 3651-60.

LaFramboise, T., B. A. Weir, X. Zhao, R. Beroukhim, C. Li, D. Harrington, W. R. Sellers & M. Meyerson, 2005. Allele-specific amplification in cancer revealed by SNP array analysis. *PLoS Comput Biol,* 1(6), e65.

Lai, W. R., M. D. Johnson, R. Kucherlapati & P. J. Park, 2005. Comparative analysis of algorithms for identifying amplifications and deletions in array CGH data. *Bioinformatics*.

Lasken, R. S. & M. Egholm, 2003. Whole genome amplification: abundant supplies of DNA from precious samples or clinical specimens. *Trends Biotechnol,* 21(12), 531-5.

Lieu, P. T., P. Jozsi, P. Gilles & T. Peterson, 2005. Development of a DNA-Labeling System for Array-Based Comparative Genomic Hybridization. *J Biomol Tech,* 16(2), 104-11.

Lisitsyn, N. & M. Wigler, 1993. Cloning the differences between two complex genomes. *Science,* 259(5097), 946-51.

Lo, K. C., L. C. Stein, J. A. Panzarella, J. K. Cowell & L. Hawthorn, 2008. Identification of genes involved in squamous cell carcinoma of the lung using synchronized data from DNA copy number and transcript expression profiling analysis. *Lung Cancer,* 59(3), 315-31.

Lockwood, W. W., R. Chari, B. Chi & W. L. Lam, 2005. Recent Advances in Array Comparative Genomic Hybridization Technologies and their Applications in Human Genetics. *European Journal of Human Genetics,* in press.

Lucito, R., J. Healy, J. Alexander, A. Reiner, D. Esposito, M. Chi, L. Rodgers, A. Brady, J. Sebat, J. Troge, J. A. West, S. Rostan, K. C. Nguyen, S. Powers, K. Q. Ye, A. Olshen, E. Venkatraman, L. Norton & M. Wigler, 2003. Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation. *Genome Res,* 13(10), 2291-305.

Lucito, R., J. West, A. Reiner, J. Alexander, D. Esposito, B. Mishra, S. Powers, L. Norton & M. Wigler, 2000. Detecting gene copy number fluctuations in tumor cells by microarray analysis of genomic representations. *Genome Res,* 10(11), 1726-36.

Mantripragada, K. K., P. G. Buckley, C. Jarbo, U. Menzel & J. P. Dumanski, 2003. Development of NF2 gene specific, strictly sequence defined diagnostic microarray for deletion detection. *J Mol Med,* 81(7), 443-51.

Margolin, A. A., J. Greshock, T. L. Naylor, Y. Mosse, J. M. Maris, G. Bignell, A. I. Saeed, J. Quackenbush & B. L. Weber, 2005. CGHAnalyzer: a stand-alone software package for cancer genome analysis using array-based DNA copy number data. *Bioinformatics,* 21(15), 3308-11.

Marra, M. A., T. A. Kucaba, N. L. Dietrich, E. D. Green, B. Brownstein, R. K. Wilson, K. M. McDonald, L. W. Hillier, J. D. McPherson & R. H. Waterston, 1997. High throughput fingerprint analysis of large-insert clones. *Genome Res,* 7(11), 1072-84.

Marx, K. A., J. R. Allen & J. E. Hearst, 1976. Characterization of the repetitious human DNA families. *Biochim Biophys Acta,* 425(2), 129-47.

Massion, P. P., W. L. Kuo, D. Stokoe, A. B. Olshen, P. A. Treseler, K. Chin, C. Chen, D. Polikoff, A. N. Jain, D. Pinkel, D. G. Albertson, D. M. Jablons & J. W. Gray, 2002. Genomic copy number analysis of non-small cell lung cancer using array comparative genomic hybridization: implications of the phosphatidylinositol 3-kinase pathway. *Cancer Res,* 62(13), 3636-40.

Matsuzaki, H., S. Dong, H. Loi, X. Di, G. Liu, E. Hubbell, J. Law, T. Berntsen, M. Chadha, H. Hui, G. Yang, G. C. Kennedy, T. A. Webster, S. Cawley, P. S. Walsh, K. W. Jones, S. P. Fodor & R. Mei, 2004. Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. *Nat Methods,* 1(2), 109-11.

Mc Sherry, E. A., A. Mc Goldrick, E. W. Kay, A. M. Hopkins, W. M. Gallagher & P. A. Dervan, 2007. Formalin-fixed paraffin-embedded clinical tissues show spurious copy number changes in array-CGH profiles. *Clin Genet,* 72(5), 441-7.

McPherson, J. D., M. Marra, L. Hillier, R. H. Waterston, A. Chinwalla, J. Wallis, M. Sekhon, K. Wylie, E. R. Mardis, R. K. Wilson, R. Fulton, T. A. Kucaba, C. Wagner-McPherson, W. B. Barbazuk, S. G. Gregory, S. J. Humphray, L. French, R. S. Evans, G. Bethel, A. Whittaker, J. L. Holden, O. T. McCann, A. Dunham, C. Soderlund, C. E. Scott, D. R. Bentley, G. Schuler, H. C. Chen, W. Jang, E. D. Green, J. R. Idol, V. V. Maduro, K. T. Montgomery, E. Lee, A. Miller, S. Emerling, Kucherlapati, R. Gibbs, S. Scherer, J. H. Gorrell, E. Sodergren, K. Clerc-Blankenburg, P. Tabor, S. Naylor, D. Garcia, P. J. de Jong, J. J. Catanese, N. Nowak, K. Osoegawa, S. Qin, L. Rowen, A. Madan, M. Dors, L. Hood, B. Trask, C. Friedman, H. Massa, V. G. Cheung, I. R. Kirsch, T. Reid, R. Yonescu, J. Weissenbach, T. Bruls, R. Heilig, E. Branscomb, A. Olsen, N. Doggett, J. F. Cheng, T. Hawkins, R. M. Myers, J. Shang, L. Ramirez, J. Schmutz, O. Velasquez, K. Dixon, N. E. Stone, D. R. Cox, D. Haussler, W. J. Kent, T. Furey, S. Rogic, S. Kennedy, S. Jones, A. Rosenthal, G. Wen, M. Schilhabel, G. Gloeckner, G. Nyakatura, R. Siebert, B. Schlegelberger, J. Korenberg, X. N. Chen, A. Fujiyama, M. Hattori, A. Toyoda, T. Yada, H. S. Park, Y. Sakaki, N. Shimizu, S. Asakawa, K. Kawasaki, T. Sasaki, A. Shintani, A. Shimizu, K. Shibuya, J. Kudoh, S. Minoshima, J. Ramser, P. Seranski, C. Hoff, A. Poustka, R. Reinhardt & H. Lehrach, 2001. A physical map of the human genome. *Nature,* 409(6822), 934-41.

McQuain, M. K., K. Seale, J. Peek, S. Levy & F. R. Haselton, 2003. Effects of relative humidity and buffer additives on the contact printing of microarrays by quill pins. *Anal Biochem,* 320(2), 281-91.

Nessling, M., K. Richter, C. Schwaenen, P. Roerig, G. Wrobel, S. Wessendorf, B. Fritz, M. Bentz, H. P. Sinn, B. Radlwimmer & P. Lichter, 2005. Candidate genes in breast cancer revealed by microarray-based comparative genomic hybridization of archived tissue. *Cancer Res,* 65(2), 439-47.

Nymark, P., H. Wikman, S. Ruosaari, J. Hollmen, E. Vanhala, A. Karjalainen, S. Anttila & S. Knuutila, 2006. Identification of specific gene copy number changes in asbestos-related lung cancer. *Cancer Res,* 66(11), 5737-43.

Osoegawa, K., A. G. Mammoser, C. Wu, E. Frengen, C. Zeng, J. J. Catanese & P. J. de Jong, 2001. A bacterial artificial chromosome library for sequencing the complete human genome. *Genome Res,* 11(3), 483-96.

Peng, W. X., T. Shibata, H. Katoh, A. Kokubu, Y. Matsuno, H. Asamura, R. Tsuchiya, Y. Kanai, F. Hosoda, T. Sakiyama, M. Ohki, I. Imoto, J. Inazawa & S. Hirohashi, 2005. Array-based comparative genomic hybridization analysis of high-grade neuroendocrine tumors of the lung. *Cancer Sci,* 96(10), 661-7.

Petrescu, A. D., H. R. Payne, A. Boedecker, H. Chao, R. Hertz, J. Bar-Tana, F. Schroeder & A. B. Kier, 2003. Physical and functional interaction of Acyl-CoA-binding protein with hepatocyte nuclear factor-4 alpha. *J Biol Chem,* 278(51), 51813-24.

Pfeifer, G., S. Steigerwald, P. Mueller, B. Wold & A. Riggs, 1989. Genomic sequencing and methylation analysis by ligation mediated PCR. *Science,* 246, 810-3.

Pinkel, D. & D. G. Albertson, 2005a. Array comparative genomic hybridization and its applications in cancer. *Nat Genet,* 37 Suppl, S11-7.

Pinkel, D. & D. G. Albertson, 2005b. Comparative genomic hybridization. *Annu Rev Genomics Hum Genet,* 6, 331-54.

Pollack, J. R., C. M. Perou, A. A. Alizadeh, M. B. Eisen, A. Pergamenschikov, C. F. Williams, S. S. Jeffrey, D. Botstein & P. O. Brown, 1999. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nature Genet,* 23(1), 41-6.

Quackenbush, J., 2002. Microarray data normalization and transformation. *Nature Genet,* 32(Suppl), 496-501.

Redon, R., M. Rio, S. G. Gregory, R. A. Cooper, H. Fiegler, D. Sanlaville, R. Banerjee, C. Scott, P. Carr, C. Langford, V. Cormier-Daire, A. Munnich, N. P. Carter & L. Colleaux, 2005. Tiling path resolution mapping of constitutional 1p36 deletions by array-CGH: contiguous gene deletion or "deletion with positional effect" syndrome? *J Med Genet,* 42(2), 166-71.

Rekhter, M. D. & J. Chen, 2001. Molecular analysis of complex tissues is facilitated by laser capture microdissection: critical role of upstream tissue processing. *Cell Biochem Biophys,* 35(1), 103-13.

Renz, M. & C. Kurz, 1984. A colorimetric method for DNA hybridization. *Nucleic Acids Res,* 12(8), 3435-44.

Richter, A., C. Schwager, S. Hentze, W. Ansorge, M. W. Hentze & M. Muckenthaler, 2002. Comparison of fluorescent tag DNA labeling methods used for expression analysis by DNA microarrays. *Biotechniques,* 33(3), 620-8, 30.

Roerig, P., M. Nessling, B. Radlwimmer, S. Joos, G. Wrobel, C. Schwaenen, G. Reifenberger & P. Lichter, 2005. Molecular classification of human gliomas using matrix-based comparative genomic hybridization. *Int J Cancer,* 117(1), 95-103.

Sambrook, J., E. F. Fritsch & T. Maniatis, (1989). Molecular Cloning: A laboratory manual, Cold Spring Harbor: Cold Spring Harbor Laboratory.

Schrock, E., S. du Manoir, T. Veldman, B. Schoell, J. Wienberg, M. A. Ferguson-Smith, Y. Ning, D. H. Ledbetter, I. Bar-Am, D. Soenksen, Y. Garini & T. Ried, 1996. Multicolor spectral karyotyping of human chromosomes. *Science,* 273(5274), 494-7.

Schwaenen, C., M. Nessling & S. Wessendorf, 2004. Automated array-based genomic profiling in chronic lymphocytic leukemia: development of a clinical tool and discovery of recurrent genomic alterations. *Proc Natl Acad Sci U S A,* 101(4), 1039-44.

Sebat, J., B. Lakshmi, J. Troge, J. Alexander, J. Young, P. Lundin, S. Maner, H. Massa, M. Walker, M. Chi, N. Navin, R. Lucito, J. Healy, J. Hicks, K. Ye, A. Reiner, T. C. Gilliam, B. Trask, N. Patterson, A. Zetterberg & M. Wigler, 2004. Large-scale copy number polymorphism in the human genome. *Science,* 305(5683), 525-8.

Sharp, A. J., D. P. Locke, S. D. McGrath, Z. Cheng, J. A. Bailey, R. U. Vallente, L. M. Pertz, R. A. Clark, S. Schwartz, R. Segraves, V. V. Oseroff, D. G. Albertson, D. Pinkel & E. E. Eichler, 2005. Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet,* 77(1), 78-88.

Shibata, T., S. Uryu, A. Kokubu, F. Hosoda, M. Ohki, T. Sakiyama, Y. Matsuno, R. Tsuchiya, Y. Kanai, T. Kondo, I. Imoto, J. Inazawa & S. Hirohashi, 2005. Genetic classification of lung adenocarcinoma based on array-based comparative genomic hybridization analysis: its association with clinicopathologic features. *Clin Cancer Res,* 11(17), 6177-85.

Simard, C., R. Lemieux & S. Cote, 2001. Urea substitutes toxic formamide as destabilizing agent in nucleic acid hybridizations with RNA probes. *Electrophoresis,* 22(13), 2679-83.

Siwoski, A., A. Ishkanian, C. Garnis, L. Zhang, M. Rosin & W. L. Lam, 2002. An efficient method for the assessment of DNA quality of archival microdissected specimens. *Modern Path,* 15(8), 889-92.

Snijders, A. M., N. Nowak, R. Segraves, S. Blackwood, N. Brown, J. Conroy, G. Hamilton, A. K. Hindle, B. Huey, K. Kimura, S. Law, K. Myambo, J. Palmer, B. Ylstra, J. P. Yue, J. W. Gray, A. N. Jain, D. Pinkel & D. G. Albertson, 2001. Assembly of microarrays for genome-wide measurement of DNA copy number. *Nat Genet,* 29(3), 263-4.

Solinas-Toldo, S., S. Lampel, S. Stilgenbauer, J. Nickolenko, A. Benner, H. Dohner, T. Cremer & P. Lichter, 1997. Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes, Chromosomes & Cancer.,* 20(4), 399-407.

Solomon, N. M., S. A. Ross, T. Morgan, J. L. Belsky, F. A. Hol, P. S. Karnes, N. J. Hopwood, S. E. Myers, A. S. Tan, G. L. Warne, S. M. Forrest & P. Q. Thomas, 2004. Array comparative genomic hybridisation analysis of boys with X linked hypopituitarism identifies a 3.9 Mb duplicated critical region at Xq27 containing SOX3. *J Med Genet,* 41(9), 669-78.

Southern, E. M., 1975. Long range periodicities in mouse satellite DNA. *J. Mol. Biol.,* 94, 51-69.

Telenius, H., N. P. Carter, C. E. Bebb, M. Nordenskjold, B. A. Ponder & A. Tunnacliffe, 1992. Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. *Genomics,* 13(3), 718-25.

Ting, A. C., S. F. Lee & K. Wang, 2003. An easy setup for double slide microarray hybridization. *Biotechniques,* 35(4), 808-10.

Tonon, G., K. K. Wong, G. Maulik, C. Brennan, B. Feng, Y. Zhang, D. B. Khatry, A. Protopopov, M. J. You, A. J. Aguirre, E. S. Martin, Z. Yang, H. Ji, L. Chin & R. A. Depinho, 2005. High-resolution genomic profiles of human lung cancer. *Proc Natl Acad Sci U S A,* 102(27), 9625-30.

Travis, W. D., 2002. Pathology of lung cancer. *Clin Chest Med,* 23(1), 65-81, viii.

Tsubosa, Y., H. Sugihara, K. Mukaisho, S. Kamitani, D. F. Peng, Z. Q. Ling, T. Tani & T. Hattori, 2005. Effects of degenerate oligonucleotide-primed polymerase chain reaction amplification and labeling methods on the sensitivity and specificity of metaphase- and array-based comparative genomic hybridization. *Cancer Genet Cytogenet,* 158(2), 156-66.

van Duin, M., R. van Marion, J. E. Watson, P. L. Paris, A. Lapuk, N. Brown, V. V. Oseroff, D. G. Albertson, D. Pinkel, P. de Jong, E. P. Nacheva, W. Dinjens, H. van Dekken & C. Collins, 2005. Construction and application of a full-coverage, high-resolution, human chromosome 8q genomic microarray for comparative genomic hybridization. *Cytometry A,* 63(1), 10-9.

Wahl, G. M., M. Stern & G. R. Stark, 1979. Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl-paper and rapid hybridization by using dextran sulfate. *Proc Natl Acad Sci U S A,* 76(8), 3683-7.

Watson, S. K., R. J. DeLeeuw, A. S. Ishkanian, C. A. Malloff & W. L. Lam, 2004. Methods for high throughput validation of amplified fragment pools of BAC DNA for constructing high resolution CGH arrays. *BMC Genomics,* 5(1), 6.

Weir, B. A., M. S. Woo, G. Getz, S. Perner, L. Ding, R. Beroukhim, W. M. Lin, M. A. Province, A. Kraja, L. A. Johnson, K. Shah, M. Sato, R. K. Thomas, J. A. Barletta, I. B. Borecki, S. Broderick, A. C. Chang, D. Y. Chiang, L. R. Chirieac, J. Cho, Y. Fujii, A. F. Gazdar, T. Giordano, H. Greulich, M. Hanna, B. E. Johnson, M. G. Kris, A. Lash, L. Lin, N. Lindeman, E. R. Mardis, J. D. McPherson, J. D. Minna, M. B. Morgan, M. Nadel, M. B. Orringer, J. R. Osborne, B. Ozenberger, A. H. Ramos, J. Robinson, J. A. Roth, V. Rusch, H. Sasaki, F. Shepherd, C. Sougnez, M. R. Spitz, M. S. Tsao, D. Twomey, R. G. Verhaak, G. M. Weinstock, D. A. Wheeler, W. Winckler, A. Yoshizawa, S. Yu, M. F. Zakowski, Q. Zhang, D. G. Beer, Wistuba, II, M. A. Watson, L. A. Garraway, M. Ladanyi, W. D. Travis, W. Pao, M. A. Rubin, S. B. Gabriel, R. A. Gibbs, H. E. Varmus, R. K. Wilson, E. S. Lander & M. Meyerson, 2007. Characterizing the cancer genome in lung adenocarcinoma. *Nature,* 450(7171), 893-8.

Wong, K. K., R. J. deLeeuw, N. S. Dosanjh, L. R. Kimm, Z. Cheng, D. E. Horsman, C. MacAulay, R. T. Ng, C. J. Brown, E. E. Eichler & W. L. Lam, 2007. A comprehensive analysis of common copy-number variations in the human genome. *Am J Hum Genet,* 80(1), 91-104.

Xiang, C. C., O. A. Kozhich, M. Chen, J. M. Inman, Q. N. Phan, Y. Chen & M. J. Brownstein, 2002. Amine-modified random primers to label probes for DNA microarrays. *Nat Biotechnol,* 20(7), 738-42.

Yu, H., J. Chao, D. Patek, R. Mujumdar, S. Mujumdar & A. S. Waggoner, 1994. Cyanine dye dUTP analogs for enzymatic labeling of DNA probes. *Nucleic Acids Res,* 22(15), 3226-32.

Yu, J., M. I. Othman, R. Farjo, S. Zareparsi, S. P. MacNee, S. Yoshida & A. Swaroop, 2002. Evaluation and optimization of procedures for target labeling and hybridization of cDNA microarrays. *Mol Vis,* 8, 130-7.

Zafarana, G., B. Grygalewicz, A. J. Gillis, L. E. Vissers, W. van de Vliet, R. J. van Gurp, H. Stoop, M. Debiec-Rychter, J. W. Oosterhuis, A. G. van Kessel, E. F. Schoenmakers, L. H. Looijenga & J. A. Veltman, 2003. 12p-amplicon structure analysis in testicular germ cell tumors of adolescents and adults by array CGH. *Oncogene,* 22(48), 7695-701.

Zhao, X., B. A. Weir, T. LaFramboise, M. Lin, R. Beroukhim, L. Garraway, J. Beheshti, J. C. Lee, K. Naoki, W. G. Richards, D. Sugarbaker, F. Chen, M. A. Rubin, P. A. Janne, L. Girard, J. Minna, D. Christiani, C. Li, W. R. Sellers & M. Meyerson, 2005. Homozygous deletions and chromosome amplifications in human lung carcinomas revealed by single nucleotide polymorphism array analysis. *Cancer Res,* 65(13), 5561-70.

3. CHROMOSOME 5P ABERRATIONS ARE EARLY EVENTS IN LUNG CANCER: IMPLICATION OF GLIAL CELL LINE-DERIVED NEUROTROPHIC FACTOR IN DISEASE PROGRESSION[9]

3.1. Introduction

Lung cancer is the most commonly diagnosed cancer worldwide, with over 170 000 new cases diagnosed last year in the United States alone (Jemal et al. 2003). The vast majority of cases are caused by smoking tobacco, with non-small-cell lung cancer (NSCLC) accounting for approximately 80% of cases (Travis et al. 1999). Squamous cell carcinoma (SqCC), the squamous subtype of NSCLC, progresses through histopathological stages from various degrees of hyperplasia and dysplasia, to carcinoma *in situ* (CIS), to invasive carcinoma, and finally to metastasis (Garnis et al. 2004a; Minna et al. 2002; Osada & Takahashi 2002). Allelotype analysis of NSCLC and cell line derivatives has yielded much insight into genetic alterations in lung cancer and gene expression profiling studies have contributed to disease subclassification. However, there are few reports on early-stage lesions, mainly due to the difficulty in obtaining such specimens. Studies of chromosome arm 3p, among others, have provided evidence that cumulative molecular alterations do accompany the progressive morphological changes that occur during the multi-stage development of lung SqCC (Minna et al. 2002; Wistuba et al. 2000; Zabarovsky et al. 2002).

While genetic alterations in tumors are common, examination of tissues earlier in tumorigenesis is more likely to identify causal events leading to tumor progression. These important early alterations may be masked by the complex pattern of genetic alterations often associated with genetic instability in the later stages of disease. Molecular cytogenetic studies have shown that chromosomal aberrations occur on the short arm of chromosome 5 (5p) in all major lung tumor types (Balsara et al. 1997; Luk et al. 2001; Ried et al. 1994; Ullmann et al. 1998). We have recently developed a 5p-specific bacterial artificial chromosome (BAC) genomic array for use in array

[9] A version of this chapter has been published: Garnis C, Davies JJ, Buys TP, Tsao MS, MacAulay C, Lam S, Lam WL. (**2005**) "Chromosome 5p aberrations are early events in lung cancer: implication of glial cell line-derived neurotrophic factor in disease progression" *Oncogene*, 24(30):4806-12.

comparative genomic hybridization (CGH) analysis. This array comprehensively covers 5p in 491 overlapping segments, allowing for high-resolution detection of copy number alterations (Coe et al. 2005). The ability to profile the entire 5p chromosome arm in one experiment – as compared to traditional marker-based techniques – is key to the analysis of minute pre-invasive CIS lesions due to the fact that these lesions are small and require microdissection which yields limited DNA.

In this study, we analyzed bronchial CIS lesions and SqCC tumors through the use of a unique 5p-specific BAC array. Alignment of array CGH profiles revealed multiple distinct regions of amplification and deletion in CIS that would otherwise appear to be masked in the later-stage tumor samples. The identification of these recurrent alterations not only led to the discovery of candidate genes but also highlights the importance of studying the genomics of early-stage lung cancer.

3.2. Results

3.2.1. 5p array CGH analysis of lung CIS and tumors

5p aberrations have been shown to be frequent in NSCLC. To investigate the role of 5p in the progression of CIS (pre-invasive) to tumors (invasive), we compared eight microdissected bronchial CIS lesions and nine SqCC tumor samples for genetic alterations using a chromosome arm-specific array containing 491 overlapping BAC clones that spans 5p from centromere (p11) to telomere (p15.33) (Coe et al. 2005). Signal intensity ratios for each triplicate-spotted BAC (giving a total of >25 000 data points across all 17 samples) were calculated and displayed as $\log_2$ plots using SeeGH software (Chi et al. 2004). Figure 3.1A gives examples of 5p SeeGH karyograms illustrating a profile with minimal copy number changes (T5850), a profile with whole arm amplification (T11773), and profiles with multiple segmental gains and losses (T10999, T8611). Alterations were detected in nearly all of the samples analyzed: in seven of the eight CIS samples and all of the tumors. Figure 3.1B shows the alignment of all 17 profiles, displayed in Java TreeView with colorimetric representation. Four tumors and one CIS (T11278, T3010, T8611, T11773, C125) showed amplifications spanning nearly the entire chromosome arm (Figure 3.1B). Notably, smaller and more

distinct regions of gain and loss were apparent in the CIS profiles, for example, the amplified regions A3 and A4 within 5p15.2. Furthermore, recurrent deletions were prominent in CIS but undetected in tumors, suggesting the possibility that deletions could be masked in tumors by later genetic events (Figure 3.1B).

Our data revealed broader alterations in tumors relative to those seen in CIS samples. Array CGH profile alignment enabled identification of six recurrent regions of amplification amongst the tumor and CIS samples (A1–A6) and three recurring regions of deletion within the CIS samples (D1–D3) (Figure 3.1B, Table 3.1). Due to the comprehensive nature of the 5p array, we observed submegabase changes which account for seven of the nine recurrent regions of alteration. These small alterations would likely have escaped detection through the use of traditional marker-based techniques.

3.2.2. Novel amplifications defined in CIS lesions

To establish a threshold, normal male DNA versus normal male DNA hybridizations were used to determine the background experimental noise at three standard deviations (Garnis et al. 2003). This placed our threshold at a 70.13 $\log_2$ raw ratio, which was used to define regions of genetic alteration. In addition, for a region to be defined as altered, it had to involve consecutive overlapping BAC clones. To increase the likelihood of identifying causal genetic events, we applied stringent selection criteria requiring the presence of alterations in at least half of the CIS samples to define a recurrent region. Strikingly, of the six recurrent amplified regions defined, five (Regions A2–A6) have not been previously reported in SqCC. The previously reported amplification at 5p15 (Bryce et al. 2000; Saretzki et al. 2002), which contains the well-characterized *Human Telomerase Reverse Transcriptase* gene (*hTERT*), falls within Region A1. However, the other locus known to be amplified in tumors (5p13.2), containing *S-phase kinase-associated protein 2* (*Skp2*), although detected in the array CGH tumor profiles, did not qualify as a region of interest in our study due to the absence of amplification in the CIS samples at this locus (Zhu et al. 2004). Remarkably, with the exception of Region A1, all other amplifications were <1 megabasepair (Mbp) in size (Table 3.1). Regions A2, A3, A4, and A6 contain only a

single annotated gene, while A5 contains only two known genes. The boundaries of these narrow regions are listed in Table 3.1. Of the four genes in single gene regions, *Catenin Delta-2* (*CTNND2*) has been shown to be over-expressed in prostate cancer (Burger et al. 2002) and *Steroid 5-Alpha-Reductase* (*SRD5A1*), which catalyses the conversion of testosterone into the more potent androgen dihydrotestosterone (Harris et al. 1992), has been shown to be over-expressed in breast cancer (Wiebe & Lewis 2003). The *Triple Functional Domain* gene (*TRIO*) in Region A4 (Figure 3.1C) and the *Glial Cell Line-Derived Neurotrophic Factor* gene (*GDNF*) in Region A6 (Figure 3.1D), however, have not been directly implicated in cancer and were further investigated in this study (see below).

3.2.3. Novel deletions defined by CIS lesions

Previous chromosomal CGH studies have identified 5p amplification as a common event in NSCLC; however, copy number reduction is rarely described. Studies using microsatellite markers have revealed loss of heterozygosity (LOH) at the centromeric end of 5p (Wieland & Bohm 1994; Wieland et al. 1996). Our analysis of the tumor profiles is consistent with previous reports in that we did not detect recurrent regions of deletion in tumor samples. To our knowledge, specific recurrent regions of deletion have not been defined within 5p in NSCLC. However, when we aligned the eight CIS profiles, three distinct recurrent regions of deletion emerged (Figure 3.1B). Recurrent deletions were defined by their presence in at least four of the eight samples. Two of the three regions were <1 Mbp in size (Table 3.1). The 1.1 Mbp Region D1 contains only one known gene – *Cadherin 12* (*CDH12*) – while no genes have been annotated within Regions D2 and D3. These observations reinforce the value in analyzing pre-invasive disease, as it has led to the identification of a small deletion that contains a single gene candidate.

3.2.4. Genomic and gene expression analysis of TRIO

We have identified seven candidate genes that fall within small regions containing only one or two genes (Table 3.1). Since a detailed analysis of all the candidate genes is beyond the scope of this study, we focused on two genes: *TRIO* and *GDNF*. Figure

3.2A shows a close-up image of SeeGH karyograms for the four CIS profiles that are amplified at the *TRIO* locus. CIS samples (C59, C127, and C60) show discrete amplification at this region, whereas C125 shows a larger amplification that extends from BAC RP11-20B15 to the centromere. Amplification of the *TRIO* region is maintained in tumors (Figure 3.1C). Next, we investigated *TRIO* for differential expression by real-time PCR in eight paired normal and SqCC samples. We observed over-expression of *TRIO* in six of the eight pairs ($P < 0.05$) (Figure 3.2B). This expression pattern is confirmed in an additional panel of 13 paired normal and tumor SqCC samples by semi-quantitative reverse transcriptase-PCR, showing significant over-expression of the *TRIO* gene (Figure 3.2C, D). The concordance of copy number increase with over-expression in tumor samples implicates a role for *TRIO* in lung SqCC, while its frequent amplification in pre-invasive lesions suggests its early involvement in tumorigenesis. The TRIO protein contains three functional domains: a serine/threonine kinase domain and two guanine nucleotide exchange factor domains for the family of Rho-like GTPases, specific for Rac1 and RhoA, respectively. These functional domains suggest that this enzyme may play a key role in several signaling pathways that control cell proliferation (Debant et al. 1996). Recent reports have implicated *TRIO* in tumorigenesis (Coe et al. 2005; Zheng et al. 2004).

3.2.5. Amplification and over-expression of GDNF in pre-invasive lesions

Similar to *TRIO*, the *GDNF* region is amplified in CIS samples as well as in tumor samples (Figure 3.1D). In Figure 3.3a the amplified region containing *GDNF* is highlighted in five CIS profiles. Real-time PCR analysis showed striking over-expression of *GDNF* in five of the eight paired normal and tumor RNA samples ($P < 0.05$) (Figure 3.3B). The absence of *GDNF* expression in normal adult lung tissue agrees with a previous observation that *GDNF* is expressed in fetal but absent in post-natal lung tissue (Fromont-Hankard et al. 2002). *GDNF* expression in tumors suggests its potential involvement in tumorigenesis.

To further evaluate the tumor-specific expression of *GDNF* at the protein level, immunohistochemical analysis was performed. A monoclonal antibody specific

for GDNF showed conspicuous membrane staining with less intense cytoplasmic staining (Figure 3.3C). Staining was not detected in normal bronchial epithelium. We further analyzed a panel of samples which included hyperplasia, CIS, and tumor samples. Immunostaining was not evident in hyperplasia but was detected in CIS samples and invasive tumors. These data suggest expression of *GDNF* in early histological lung cancer stages. This is further supported by the fact GDNF is the ligand for the *RET* proto-oncogene, which is known to transduce signals for cell growth and differentiation (Takahashi 2001).

3.3. Discussion

The identification of early-stage genetic events that drive the progression of squamous cell lung cancer requires the study of pre-invasive lesions. Our data suggest the possibility that early causal events in tumorigenesis may be masked by later-stage genomic instability. The ability of the CIS profiles to narrow our focus within the tumor profiles demonstrates the importance of studying early-stage disease. In the past, such studies have been limited by the rarity and minute size of early-stage specimens. The development of the 5p genomic array has facilitated comprehensive analysis of the 5p chromosome arm for these precious samples. Consequently, we have identified nine distinct regions of alteration across 5p, of which eight are novel and seven are submegabase in size. These small alterations would likely have escaped detection through the use of limited resolution conventional marker-based techniques. This study has shown that the analysis of CIS samples with a comprehensive tiling path array is arguably one of the most effective ways to focus attention on those regions of the chromosome arm that contain genes potentially critical to disease progression.
We have assayed the expression levels of two of the candidate genes that fall within amplified regions, *TRIO* and *GDNF*, and found that both are significantly over-expressed in a panel of tumors compared with their matched normal RNA samples. More significantly, immunohistochemical analysis of GDNF, a ligand for RET, not only showed tumor-specific staining but was also present in pre-invasive stage specimens. Interestingly, *GDNF* expression is normally restricted to the fetal lung and is involved in

lung development (Fromont-Hankard et al. 2002). Our findings suggest that reactivation of this developmental gene may contribute to lung tumorigenesis at an early stage.

3.4. Materials and methods

3.4.1. Sample procurement

Formalin-fixed paraffin-embedded lung CIS samples used in array CGH analysis were collected by fluorescent bronchoscopy as part of the Lung Health Study at the British Columbia Cancer Agency (Lam et al. 2000; McWilliams et al. 2002). Both fixed and frozen tissues were obtained from the archive of St Paul's Hospital, Vancouver and the Ontario Cancer Institute, Toronto. Hematoxylin and eosin-stained sections for each sample were graded and microdissected under the guidance of a lung pathologist. DNA extraction was previously described (Siwoski et al. 2002) and RNA was isolated using Trizol reagents (Invitrogen, Burlington, ON, USA). Samples were collected following approval by the Research Ethics Boards of the respective institutions.

3.4.2. Array CGH

The establishment of the tiling set of clones spanning 5p and genomic array construction is described in Figure 3.1 legend. Array hybridizations were performed as described previously (Garnis et al. 2003; Garnis et al. 2004b). To compare multiple profiles, we used Java TreeView version 1.0.3 to generate a colored gene copy number matrix (http://jtreeview.sourceforge.net).

3.4.3. Gene expression analysis

RNA for reverse-transcriptase PCR was extracted from frozen tissue sections using Trizol reagent (Invitrogen, Burlington, ON, USA). cDNA was synthesized using the Superscript II RNAse H reverse-transcriptase system (Invitrogen). For semi-quantitative RT–PCR, expression levels were determined by gene-specific PCR. PCR cycle conditions were as follows: one cycle of 95°C, 1 min; 30–35 cycles of (95°C, 30 s; 55°C for *β-actin* and 60°C for *TRIO* for 30 s; 72°C, 1 min) and a 10 min extension at 72°C. PCR products were resolved by polyacrylamide gel electrophoresis, imaged by SYBR green staining (Roche, Laval, Que, Canada) on a Molecular Dynamics Storm

Phosphoimager model 860, and quantified using ImageQuant software (Molecular Dynamics, Piscataway, NJ, USA). The genes assayed were hypothesized to be over-expressed since they reside in amplicons. Therefore, a one-tailed Wilcoxon matched pairs signed ranks test was used to determine if over-expression of these genes was significant in sets of matched tumor and normal lung samples.

Figure 3.1 – 5p segmental copy number alterations. The physical map of the human genome (McPherson et al. 2001) and the UCSC genome browser (Kent et al. 2002) were used to facilitate the selection of a clone subset from the whole genome BAC re-array set (Krzywinski et al. 2004). This represents minimal overlapping coverage of the short arm of chromosome 5. We have made this clone list publicly available at http://www.bccrc.ca/cg/ArrayCGH_Group.html. Linker-mediated PCR (LM-PCR) of BAC DNA samples was performed to obtain sufficient concentration for spotting (Watson et al. 2004). This subset consists of 491 fingerprint-verified BAC clones spanning the >50Mbp arm of chromosome 5 from 5p11 to 5p15.33. In addition to the 5p clones, 96 random loci scattered throughout the human genome were included on the array as internal controls. LM-PCR amplified human genomic DNA spots were also added to the array in order to facilitate normalization of our signal intensities as described previously (Coe et al. 2005; Garnis et al. 2003). Array CGH profiling was performed by co-hybridizing 100 ng of reference and sample DNA labeled with cyanine 5 and cyanine 3 dCTPs, respectively. The labeling, hybridization, and imaging protocols are described previously (Garnis et al. 2003). Arrays were pre-hybridized at 42°C with DIG Easy hybridization solution (Roche, Mississauga, ON, USA) containing 1% BSA and 2 mg/ml sheared herring sperm DNA. Denatured probes in hybridization buffer containing 6 mg/ml yeast tRNA were applied to the array and hybridized at 42°C for 36 h. Arrays were washed repeatedly with 0.1x SSC/0.1% SDS at room temperature. Hybridized arrays were imaged using a CCD-based imaging system and analyzed using the SoftWorx array analysis program (ArrayWorx eAuto, API, Issaquah, WA, USA). Normalization was performed as described previously (Garnis et al. 2003; Garnis et al. 2004b). Normalized $\log_2$ signal intensity ratios were plotted using SeeGH software, version 1.7 (Chi et al. 2004). A $\log_2$ signal ratio of zero represents equivalent copy number between the sample and the reference DNA. Clones with standard deviations among the triplicate spots >0.075 or a signal-to-noise ratio <20 were disqualified from further analysis. A. 5p array CGH profiles showing normal copy number (T5850), whole arm amplification (T11773), and multiple segmental copy number changes (T10999, T8611). 5p cytoband pattern is to the left. Vertical lines

denote $\log_2$ signal ratios from -1 to 1 with copy number increases to the right (red lines) and decreases to the left (green lines) of zero (purple line). Each black dot represents a single BAC clone. B. Colorimetric representation of 5p array CGH data viewed by Java TreeView. To compare multiple profiles, we used Java TreeView version 1.0.3 to generate a colored gene copy number matrix (http://jtreeview.sourceforge.net). Intensities of red and green coloration indicate an increased or decreased $\log_2$ signal ratio for each clone, respectively. Gray coloration indicates clones discarded due to high standard deviations (>0.075) or signal-to-noise ratios. Each column represents a separate array CGH profile. Each row corresponds to a BAC clone and each column represents a CIS or tumor sample (sample ID at top). 5p cytoband pattern is to the left. Recurrent amplifications are denoted by vertical blue lines to the right and deletions to the left. The close-up views of regions A4 and A6 are shown to the right with the genes of interest represented by black lines. C. Magnification of the A4 region containing the *TRIO* gene. D. Magnification of the A6 region containing the *GDNF* gene.

Figure 3.1

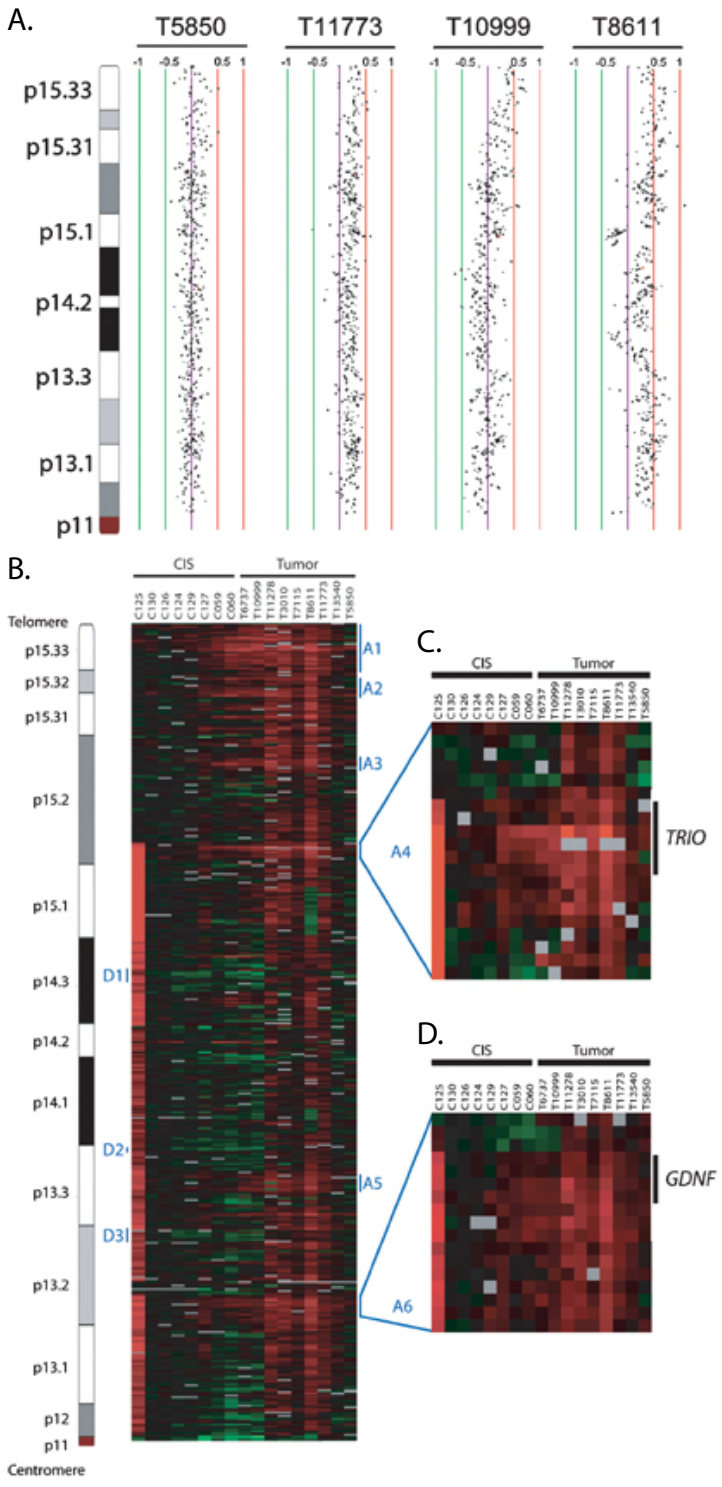

A.
T5850
T11773
T10999
T8611
p15.33
p15.31
p15.1
p14.2
p13.3
p13.1
p11
B.
CIS
Tumor
Telomere
p15.33
p15.32
p15.31
p15.2
p15.1
p14.3
p14.2
p14.1
p13.3
p13.2
p13.1
p12
p11
Centromere
A1
A2
A3
A4
A5
A6
D1
D2
D3
C.
TRIO
D.
GDNF

Table 3.1 – Recurrent minimal regions of genetic alterations on 5p.

Region[a]	Cytoband[b]	Estimated size (Mbp)	Base pair position[b]	Boundary BAC clones[c]	Known genes in region[b]
A1	p15.33	4.1	97,762 – 4,260,073	348B13 to 328F21	*hTERT, 13 others*
A2	p15.31–32	0.9	6,010,119 – 6,905,520	107F17 to 203A14	*SRD5A1*
A3	p15.2	0.15	11,062,828 – 11,209,193	471G19 to 293P3	*CTNND2*
A4	p15.2	0.27	14,196,287 – 14,471,241	20B15 to 611H4	*TRIO*
D1	p14.3	1.1	21,113,936 – 22,241,939	627F13 to 402L24	*CDH12*
D2	p13.3	0.38	29,217,035 – 29,592,188	784O7 to 258I24	(none)
A5	p13.3	0.22	31,243,308 – 31,464,867	705N10 to 784O19	*CDH6, RNASE3L*
D3	p13.2	0.075	34,098,642 – 34,173,623	110H4 to 15A6	(none)
A6	p13.2	0.34	37,711,349 – 38,049,921	285H4 to 695L19	*GDNF*

[a] A recurrent region is present in over 50% of the CIS array CGH profiles; regions correspond to those indicated on Figure 3.1A ('A' denotes amplification, 'D' denotes deletion).

[b] Base pair, cytoband, and known (reviewed) genes are based on the Human April 2003 Assembly at the UCSC Genome Browser, version hg15.

[c] BAC clones derive from the RP11 library.

Figure 3.2

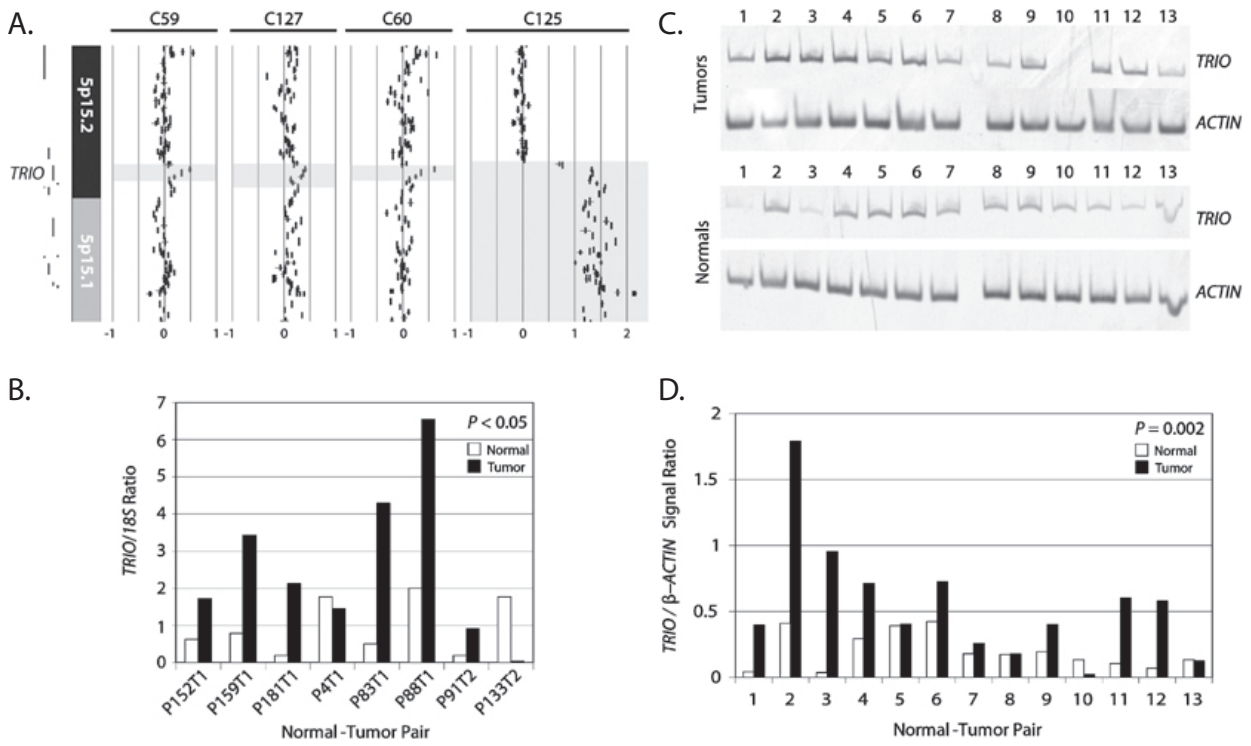
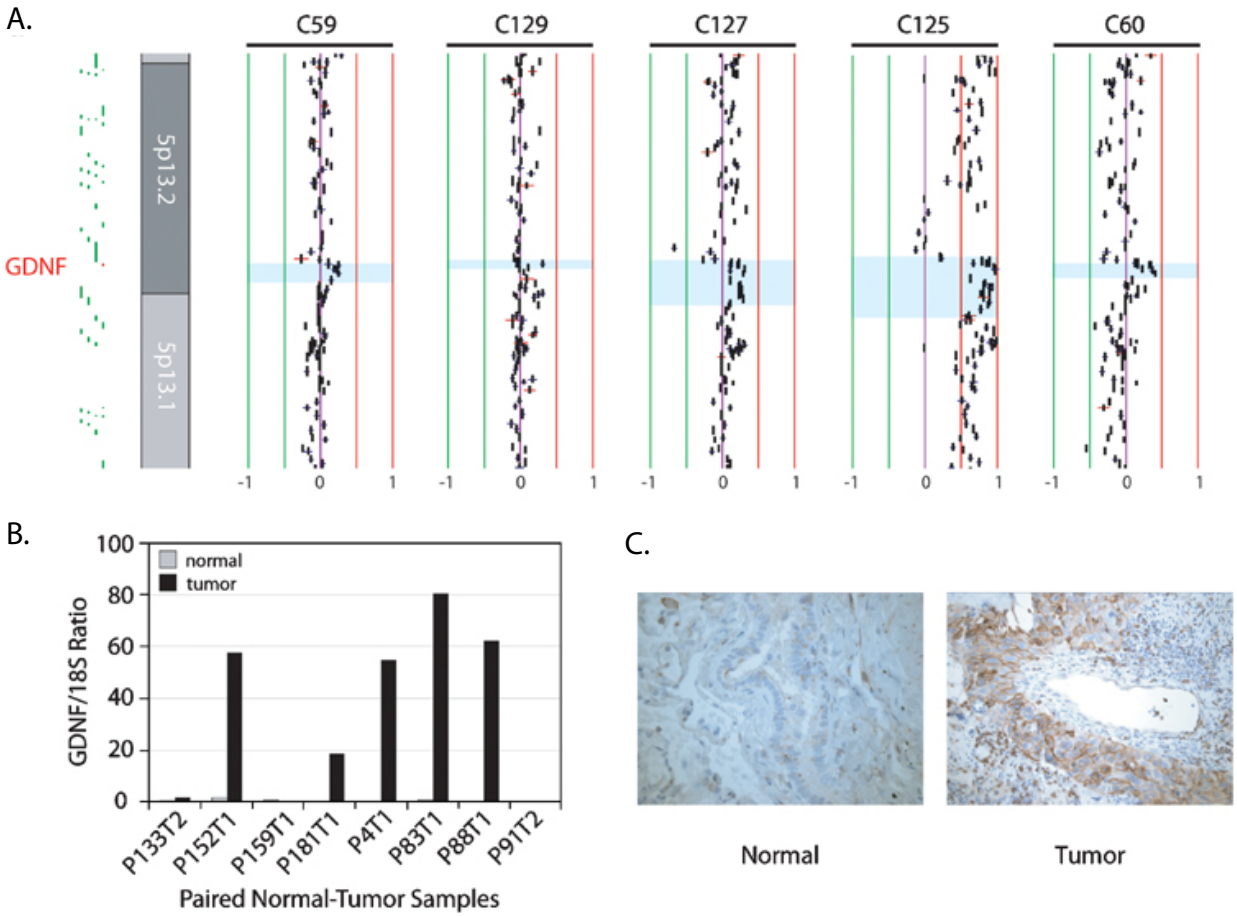


Figure 3.2 – Genomic and gene expression analysis of *TRIO*. A. SeeGH karyogram of the *TRIO* region for four CIS samples. Cytobands are indicated on the left of the profiles. Vertical lines denote $\log_2$ signal ratios with copy number increases to the right and reduction to the left of zero. Each black segment represents a single BAC clone. Amplified regions containing *TRIO* are highlighted. B. Real-time PCR expression analysis of *TRIO* in eight matched normal (gray) and tumor (black) samples. C. Reverse transcription PCR analysis of *TRIO* and *β-actin* in an additional panel of 13 matched normal and tumor samples (*TRIO* [ACTGCTGAGCACAGCTCACT, TAGAGTTTGACCTATCCAGA], *β-actin* [GATGTGGATCAGCAAGCA, GAAAGGGTGTAACGCAACT]). PCR products were resolved by polyacrylamide gel electrophoresis, stained with SYBR green, quantified using the Storm Phosphoimager, and summarized in D.

Figure 3.3 – Genomic, gene expression, and immunochemical analysis of GDNF. A. SeeGH karyogram of the GDNF region for five CIS samples. Cytobands are indicated on the left of the profiles. Vertical lines denote log2 signal ratios with copy number increases to the right (red lines) and reduction to the left (green lines) of zero (purple line). Each black segment represents a single BAC clone. Amplified region containing GDNF is highlighted. B. Real-time PCR expression analysis of GDNF in eight matched normal (gray) and tumor (black) samples. For real-time PCR analysis, mRNA expression levels were quantitatively assessed by reverse transcription-PCR using the ABI PRISM 7700 Sequence Detection System, as described previously (Wang et al., 2002). To avoid amplification of contaminating genomic DNA sequences, primers were designed to span two adjacent exons (CACTGACTTGGGTCTGGGCTAT and GTCTCAGCTGCATCGCAAG). All assays were carried out using duplicate samples of each RT product. The sample-to-sample variation of RNA/cDNA quantity was normalized using the 18S ribosomal RNA as the reference gene (Wang et al. 2002). C. Immunostaining of squamous cell lung tumor tissue sections with anti-GDNF monoclonal antibody (Santa Cruz Biotech).

Figure 3.3


A.
C59
C129
C127
C125
C60
GDNF
5p13.2
5p13.1
-1
0
1
B.
normal
tumor
GDNF/18S Ratio
100
80
60
40
20
0
P133T2
P152T1
P159T1
P181T1
P4T1
P83T1
P88T1
P91T2
Paired Normal-Tumor Samples
C.
Normal
Tumor

3.5. References

Balsara, B. R., G. Sonoda, S. du Manoir, J. M. Siegfried, E. Gabrielson & J. R. Testa, 1997. Comparative genomic hybridization analysis detects frequent, often high-level, overrepresentation of DNA sequences at 3q, 5p, 7p, and 8q in human non-small cell lung carcinomas. *Cancer Res,* 57(11), 2116-20.

Bryce, L. A., N. Morrison, S. F. Hoare, S. Muir & W. N. Keith, 2000. Mapping of the gene for the human telomerase reverse transcriptase, hTERT, to chromosome 5p15.33 by fluorescence in situ hybridization. *Neoplasia,* 2(3), 197-201.

Burger, M. J., M. A. Tebay, P. A. Keith, H. M. Samaratunga, J. Clements, M. F. Lavin & R. A. Gardiner, 2002. Expression analysis of delta-catenin and prostate-specific membrane antigen: their potential as diagnostic markers for prostate cancer. *Int J Cancer,* 100(2), 228-37.

Chi, B., R. J. deLeeuw, B. P. Coe, C. MacAulay & W. L. Lam, 2004. SeeGH - A software tool for visualization of whole genome array comparative genomic hybridization data. *BMC Bioinformatics,* 5(1), 13.

Coe, B. P., L. J. Henderson, C. Garnis, M. S. Tsao, A. F. Gazdar, J. Minna, S. Lam, C. Macaulay & W. L. Lam, 2005. High-resolution chromosome arm 5p array CGH analysis of small cell lung carcinoma cell lines. *Genes Chromosomes Cancer,* 42(3), 308-13.

Debant, A., C. Serra-Pages, K. Seipel, S. O'Brien, M. Tang, S. H. Park & M. Streuli, 1996. The multidomain protein Trio binds the LAR transmembrane tyrosine phosphatase, contains a protein kinase domain, and has separate rac-specific and rho-specific guanine nucleotide exchange factor domains. *Proc Natl Acad Sci U S A,* 93(11), 5466-71.

Fromont-Hankard, G., P. Philippe-Chomette, A. L. Delezoide, C. Nessmann, Y. Aigrain & M. Peuchmaur, 2002. Glial cell-derived neurotrophic factor expression in normal human lung and congenital cystic adenomatoid malformation. *Arch Pathol Lab Med,* 126(4), 432-6.

Garnis, C., C. Baldwin, L. Zhang, M. P. Rosin & W. L. Lam, 2003. Use of complete coverage array comparative genomic hybridization to define copy number alterations on chromosome 3p in oral squamous cell carcinomas. *Cancer Res,* 63(24), 8582-5.

Garnis, C., T. P. Buys & W. L. Lam, 2004a. Genetic alteration and gene expression modulation during cancer progression. *Mol Cancer,* 3(1), 9.

Garnis, C., B. P. Coe, A. Ishkanian, L. Zhang, M. P. Rosin & W. L. Lam, 2004b. Novel regions of amplification on 8q distinct from the MYC locus and frequently altered in oral dysplasia and cancer. *Genes Chromosomes Cancer,* 39(1), 93-8.

Harris, G., B. Azzolina, W. Baginsky, G. Cimis, G. H. Rasmusson, R. L. Tolman, C. R. Raetz & K. Ellsworth, 1992. Identification and selective inhibition of an isozyme of steroid 5 alpha-reductase in human scalp. *Proc Natl Acad Sci U S A,* 89(22), 10787-91.

Jemal, A., T. Murray, A. Samuels, A. Ghafoor, E. Ward & M. J. Thun, 2003. Cancer statistics, 2003. *CA Cancer J Clin,* 53(1), 5-26.

Kent, W. J., C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler & D. Haussler, 2002. The human genome browser at UCSC. *Genome Res,* 12(6), 996-1006.

Krzywinski, M., I. Bosdet, D. Smailus, R. Chiu, C. Mathewson, N. Wye, S. Barber, M. Brown-John, S. Chan, S. Chand, A. Cloutier, N. Girn, D. Lee, A. Masson, M. Mayo, T. Olson, P. Pandoh, A. L. Prabhu, E. Schoenmakers, M. Tsai, D. Albertson, W. Lam, C. O. Choy, K. Osoegawa, S. Zhao, P. J. de Jong, J. Schein, S. Jones & M. A. Marra, 2004. A set of BAC clones spanning the human genome. *Nucleic Acids Res,* 32(12), 3651-60.

Lam, S., C. MacAulay, J. C. leRiche & B. Palcic, 2000. Detection and localization of early lung cancer by fluorescence bronchoscopy. *Cancer,* 89(11 Suppl), 2468-73.

Luk, C., M. S. Tsao, J. Bayani, F. Shepherd & J. A. Squire, 2001. Molecular cytogenetic analysis of non-small cell lung carcinoma by spectral karyotyping and comparative genomic hybridization. *Cancer Genet Cytogenet,* 125(2), 87-99.

McPherson, J. D., M. Marra, L. Hillier, R. H. Waterston, A. Chinwalla, J. Wallis, M. Sekhon, K. Wylie, E. R. Mardis, R. K. Wilson, R. Fulton, T. A. Kucaba, C. Wagner-McPherson, W. B. Barbazuk, S. G. Gregory, S. J. Humphray, L. French, R. S. Evans, G. Bethel, A. Whittaker, J. L. Holden, O. T. McCann, A. Dunham, C. Soderlund, C. E. Scott, D. R. Bentley, G. Schuler, H. C. Chen, W. Jang, E. D. Green, J. R. Idol, V. V. Maduro, K. T. Montgomery, E. Lee, A. Miller, S. Emerling, R. S. Kucherlapati, R. Gibbs, S. Scherer, J. H. Gorrell, E. Sodergren, K. Clerc-Blankenburg, P. Tabor, S. Naylor, D. Garcia, P. J. de Jong, J. J. Catanese, N. Nowak, K. Osoegawa, S. Qin, L. Rowen, A. Madan, M. Dors, L. Hood, B. Trask, C. Friedman, H. Massa, V. G. Cheung, I. R. Kirsch, T. Reid, R. Yonescu, J. Weissenbach, T. Bruls, R. Heilig, E. Branscomb, A. Olsen, N. Doggett, J. F. Cheng, T. Hawkins, R. M. Myers, J. Shang, L. Ramirez, J. Schmutz, O. Velasquez, K. Dixon, N. E. Stone, D. R. Cox, D. Haussler, W. J. Kent, T. Furey, S. Rogic, S. Kennedy, S. Jones, A. Rosenthal, G. Wen, M. Schilhabel, G. Gloeckner, G. Nyakatura, R. Siebert, B. Schlegelberger, J. Korenberg, X. N.

Chen, A. Fujiyama, M. Hattori, A. Toyoda, T. Yada, H. S. Park, Y. Sakaki, N. Shimizu, S. Asakawa, K. Kawasaki, T. Sasaki, A. Shintani, A. Shimizu, K. Shibuya, J. Kudoh, S. Minoshima, J. Ramser, P. Seranski, C. Hoff, A. Poustka, R. Reinhardt & H. Lehrach, 2001. A physical map of the human genome. *Nature,* 409(6822), 934-41.

McWilliams, A., C. MacAulay, A. F. Gazdar & S. Lam, 2002. Innovative molecular and imaging approaches for the detection of lung cancer and its precursor lesions. *Oncogene,* 21(45), 6949-59.

Minna, J. D., J. A. Roth & A. F. Gazdar, 2002. Focus on lung cancer. *Cancer Cell,* 1(1), 49-52.

Osada, H. & T. Takahashi, 2002. Genetic alterations of multiple tumor suppressors and oncogenes in the carcinogenesis and progression of lung cancer. *Oncogene,* 21(48), 7421-34.

Ried, T., I. Petersen, H. Holtgreve-Grez, M. R. Speicher, E. Schrock, S. du Manoir & T. Cremer, 1994. Mapping of multiple DNA gains and losses in primary small cell lung carcinomas by comparative genomic hybridization. *Cancer Res,* 54(7), 1801-6.

Saretzki, G., S. Petersen, I. Petersen, K. Kolble & T. von Zglinicki, 2002. hTERT gene dosage correlates with telomerase activity in human lung cancer cell lines. *Cancer Lett,* 176(1), 81-91.

Siwoski, A., A. Ishkanian, C. Garnis, L. Zhang, M. Rosin & W. L. Lam, 2002. An efficient method for the assessment of DNA quality of archival microdissected specimens. *Mod Pathol,* 15(8), 889-92.

Takahashi, M., 2001. The GDNF/RET signaling pathway and human diseases. *Cytokine Growth Factor Rev,* 12(4), 361-73.

Travis, W. D., T. V. Colby, B. Corrin, Y. Shimosato & E. Brambilla, 1999. *Histological Typing of Lung and Pleural Tumours with contributions by Pathologists from 14 Countries,* Berlin: Springer Verlag.

Ullmann, R., A. Schwendel, H. Klemen, G. Wolf, I. Petersen & H. H. Popper, 1998. Unbalanced chromosomal aberrations in neuroendocrine lung tumors as detected by comparative genomic hybridization. *Hum Pathol,* 29(10), 1145-9.

Wang, K. K., N. Liu, N. Radulovich, D. A. Wigle, M. R. Johnston, F. A. Shepherd, M. D. Minden & M. S. Tsao, 2002. Novel candidate tumor marker genes for lung adenocarcinoma. *Oncogene,* 21(49), 7598-604.

Watson, S. K., R. J. deleeuw, A. S. Ishkanian, C. A. Malloff & W. L. Lam, 2004. Methods for high throughput validation of amplified fragment pools of BAC DNA for constructing high resolution CGH arrays. *BMC Genomics,* 5(1), 6.

Wiebe, J. P. & M. J. Lewis, 2003. Activity and expression of progesterone metabolizing 5alpha-reductase, 20alpha-hydroxysteroid oxidoreductase and 3alpha(beta)-hydroxysteroid oxidoreductases in tumorigenic (MCF-7, MDA-MB-231, T-47D) and nontumorigenic (MCF-10A) human breast cancer cells. *BMC Cancer,* 3(1), 9.

Wieland, I. & M. Bohm, 1994. Frequent allelic deletion at a novel locus on chromosome 5 in human lung cancer. *Cancer Res,* 54(7), 1772-4.

Wieland, I., M. Bohm, K. C. Arden, T. Ammermuller, S. Bogatz, C. S. Viars & M. F. Rajewsky, 1996. Allelic deletion mapping on chromosome 5 in human carcinomas. *Oncogene,* 12(1), 97-102.

Wistuba, I., J. Berry, C. Behrens, A. Maitra, N. Shivapurkar, S. Milchgrub, B. Mackay, J. D. Minna & A. F. Gazdar, 2000. Molecular changes in the bronchial epithelium of patients with small cell lung cancer. *Clin Cancer Res,* 6(7), 2604-10.

Zabarovsky, E. R., M. I. Lerman & J. D. Minna, 2002. Tumor suppressor genes on chromosome 3p involved in the pathogenesis of lung and other cancers. *Oncogene.,* 21(45), 6915-35.

Zheng, M., R. Simon, M. Mirlacher, R. Maurer, T. Gasser, T. Forster, P. A. Diener, M. J. Mihatsch, G. Sauter & P. Schraml, 2004. TRIO amplification and abundant mRNA expression is associated with invasive tumor growth and rapid tumor cell proliferation in urinary bladder cancer. *Am J Pathol,* 165(1), 63-9.

Zhu, C. Q., F. H. Blackhall, M. Pintilie, P. Iyengar, N. Liu, J. Ho, T. Chomiak, D. Lau, T. Winton, F. A. Shepherd & M. S. Tsao, 2004. Skp2 gene copy number aberrations are common in non-small cell lung carcinoma, and its overexpression in tumors with ras mutation is a poor prognostic marker. *Clin Cancer Res,* 10(6), 1984-91.

4. INTEGRATIVE GENOMIC AND GENE EXPRESSION ANALYSIS OF CHROMOSOME 7 IDENTIFIED NOVEL ONCOGENE LOCI IN NON-SMALL CELL LUNG CANCER[10]

4.1. Introduction

Lung cancer is broadly classified into small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC) types, with the latter group accounting for approximately 80% of cases and having an overall 5-year survival rate of ~15% (Travis et al. 1999). NSCLC is subclassified into multiple subtypes including adenocarcinomas (AC), squamous cell carcinomas (SqCC), and large cell carcinomas (LCC). While DNA sequence mutations and epigenetic alterations have been demonstrated to drive lung cancer oncogene activation (Wilson et al. 2006), changes in gene dosage have consistently been shown to be a major driving force in lung tumors (Lockwood et al. 2008). Although previous work has investigated gene expression dysregulation in lung cancer cells, there have been very few attempts to incorporate the impact of genomic alterations on gene transcription levels (Coe & Lockwood et al. 2006; Tonon et al. 2005; Zhou et al. 2006). Integrative analysis of recurring genomic and gene expression alterations in NSCLC tumours and cell models may be necessary to yield new insight into lung cancer biology.

Multiple studies have shown that aberrations in DNA copy number on chromosome 7 occur frequently in lung cancer, and that gain of chromosome 7 has been associated with NSCLC aggressiveness (Pei et al. 2001), a finding consistent with observations from other cancer types (Garcia et al. 2003; Waldman et al. 1991; Arslantas et al. 2007). Both focal regions of DNA amplification and whole chromosome number imbalances have been observed (Balsara & Testa 2002; Panani & Roussos 2006; Zojer et al. 2000). Alteration at chromosome 7 is generally thought to arise through selection for activation of the *EGFR* oncogene at 7p12, though others have shown that amplification of the *MET* oncogene at 7q31 can also occur (Engelman et al. 2007). The

[10] A version of this chapter has been published: Campbell JM, Lockwood WW, Buys TPH, Chari R, Coe BP, Lam S, and Lam WL. (**2008**) "Integrative genomic and gene expression analysis of chromosome 7 identified novel oncogene loci in non-small cell lung cancer", *Genome*, 51(12):1032-9.

role of *EGFR* and *MET* activation has been elucidated in a subset of lung tumours (Engelman et al. 2007; Sequist et al. 2007). In a previous study, we observed that several regions of recurring genome alteration on chromosome 7 do not overlap with the *EGFR* and *MET* loci, raising the possibility of additional oncogene loci residing on this chromosome (Garnis et al. 2006). Those genes exhibiting over-expression in addition to genomic gain likely represent key "driver" alterations contributing to the cancer phenotype while those without concurrent mRNA increases are likely "passenger" alterations (Albertson 2006). In this study, through the integration of genomic and gene expression analyses of cell lines and clinical tumour samples, we have uncovered additional lung cancer oncogene candidates situated on chromosome 7.

4.2. Materials and methods

4.2.1. Cell line samples and DNA extraction

The 30 NSCLC cell lines used in this study are detailed in Table S1[11]. These cell lines were established at the National Cancer Institute (NCI-H series) and the Hamon Center for Therapeutic Oncology Research, University of Texas Southwestern Medical Center (HCC series). All cell lines were either acquired from The American Type Culture Collection or supplied by Dr. John Minna from the Hamon Center for Therapeutic Oncology Research and grown according to specifications. DNA was isolated using a standard procedure with Proteinase K digestion followed by phenol-chloroform extraction (Lockwood et al. 2007).

4.2.2. Tiling path array comparative genomic hybridization and data analysis

Segmental DNA copy number profiles were generated for each of the 30 NSCLC cell lines by whole genome tiling path aCGH as previously described (Watson et al. 2007; Shadeo & Lam 2006; Lockwood et al. 2007; Ishkanian et al. 2004). Images of the hybridized arrays were then analyzed using SoftWoRx Tracker Spot Analysis software

[11] Supplementary data for this manuscript are available on the journal website (http://genome.nrc.ca).

(Applied Precision), and systematic biases were removed from all array data files using a stepwise normalization procedure as previously described (Lockwood et al. 2008; Khojasteh et al. 2005). *SeeGH* was used to combine replicates and visualize all data as $\log_2$ ratio plots in karyograms (Chi et al. 2004; Chi B 2008). All raw array data files have been made publicly available through the System for Integrative Genomic Microarray Analysis (*SIGMA*) database, which can be accessed at http://sigma.bccrc.ca (Chari et al. 2006).

Regions of high level genomic amplification within each cell line were determined using an algorithm as previously described (Lockwood et al. 2008). Briefly, aCGH data was filtered to exclude clones with standard deviations between replicate values >0.075 and clones were identified as members of high level amplifications if its resulting $\log_2$ ratio was ≥0.8 (Lockwood et al. 2007; Choi et al. 2007; Choi et al. 2006). Moving averages of varying window sizes were used then used to identify amplicon boundaries (Lockwood et al. 2008).

4.2.3. Integration of copy number status and gene expression microarray data

Integration of gene expression and copy number data was performed as described previously (Lockwood et al. 2008). First, genes present in each amplicon were listed and those seen to be amplified two or more times in the 30 lines were identified. Copy number status (gain, loss or neutral) for each gene locus was defined using aCGH-Smooth (Jong et al. 2004) with lambda and breakpoint per chromosome settings at 6.75 and 100, respectively as previously described (Lockwood et al. 2008). Using these criteria, samples with neutral copy number (equal number of copies between tumor and normal reference DNA) for each gene of interest were defined. The algorithm described above was then used to determine if a cell line harbored amplification at the given locus. Array CGH measures relative and not absolute copy number and does not take into account changes due to ploidy. As such, alterations in gene dosage attributed to ploidy alone were not considered in this analysis. RNA expression profiles for 30 NSCLC cell lines were obtained from Gene Expression Omnibus (accession number

GSE4824) (Zhou et al. 2006). The Affymetrix gene expression microarray probe sets corresponding to these genes were then determined and probes filtered for those demonstrating a present or marginal quality score in at least 50% of the amplified lines. The copy number status for each gene was then dichotomized to neutral vs. amplified samples and gene expression data were then compared between the groups using the Mann-Whitney U test to identify those that were over-expressed in the amplified samples with a *p*-value ≤ 0.05.

4.2.4. Quantitative real time PCR expression analysis of cell line and clinical tumour samples

cDNA was synthesized from 5 µg of total RNA using an ABI High Capacity cDNA Archive Kit (Applied Biosystems, Foster City, CA, USA). An aliquot of 100 ng of cDNA was used for each real-time PCR reaction. TaqMan gene expression assays were performed using standard TaqMan reagents and protocols on the Applied Biosystems 7500 Fast Real-Time PCR System (Applied Biosystems). Gene expression Assay IDs used include: *NUDT1* (Hs00159343_m1), *EGFR* (Hs00193306_m1) and *TAF6* (Hs00425763_m1). Samples were run in triplicate and normalized against a *eukaryotic 18S rRNA* endogenous control (Hs99999901_s1). The relative fold change of the target gene in each cell line sample compared to a pooled normal lung cDNA reference sample (AM7968, Ambion, Austin, TX, USA) was performed using the $2^{-\Delta ddCt}$ method (Coe & Lockwood et al. 2006). The Mann-Whitney U-test was used to determine whether the expression of these genes was significantly different in cell lines with amplification compared to those with neutral copy number of the particular locus in question (*p*-value ≤ 0.05).

4.2.5. Analysis of publically available gene expression data for clinical lung tumours

Publically available Affymetrix gene expression data from 111 NSCLC tumours (53 SqCC and 58 AC) (GEO Accession number GSE3141) (Bild et al. 2006) and a normal human bronchial epithelial cell line (NHBE) (GEO Accession number GSE4824) (Zhou

et al. 2006) was used to further validate the level of expression of candidate genes in our study shown to be amplified and over-expressed in the frequently altered regions. MAS5 normalized NSCLC data were used to calculate fold changes in expression for each gene compared to the NHBE reference sample. Affymetrix probes showing the highest overall signal intensity in the tumour samples were chosen for the analysis. The number of samples having ≥2 fold over-expression for the gene of interest was calculated. Genes with ≥45% of NSCLC samples having a ≥2 fold change were determined to be significant.

4.2.6. Quantitative real time PCR expression analysis of clinical samples

Ten fresh-frozen lung NSCLC tumours and their corresponding matched normal lung tissue were obtained from Vancouver General Hospital, Vancouver, B.C., Canada. Microdissection of tumour cells was performed under the guidance of two lung pathologists. Total RNA was isolated and 1 µg was converted into cDNA as described above. Gene-specific qPCR was performed for *NUDT1, TAF6*, *POLR2J* (Hs00196523_m1)*, FTSJ2* (Hs00203647_m1) and *18S* rRNA as described above. *18S* normalized cycle thresholds for each of the genes were compared between the tumours and corresponding normal tissue from the individual patients to determine the fold change in expression using the $2^{-\Delta ddCt}$ method. Because these genes were hypothesized to be over-expressed owing to DNA amplification, a one-tailed Wilcoxon sign-rank test was used to determine whether expression of these genes was significantly different between matched tumour and normal samples (p-value < 0.05) (Lockwood et al. 2008).

4.3. Results and discussion

Segmental gain or amplification of chromosome 7 has typically been attributed to selection for *EGFR* or *MET* oncogene activation (Engelman et al. 2007; Sequist et al. 2007). It has recently been reported that additional regions of segmental genomic gain recur within chromosome 7 for NSCLC cell lines (Garnis et al. 2006) (Table 4.1). We have built upon this analysis, providing evidence for the importance of candidate

regions by (i) identifying high level amplification events and (ii) quantifying concurrent gene expression changes in cell lines and clinical samples.

4.3.1. High level amplification events in frequently altered regions

The amplification of a chromosome segment may highlight its importance in cancer development as amplicons are thought to represent the selection of genes that facilitate tumour growth. Therefore, to refine the recurring regions of copy number gain on chromosome 7, we first evaluated which of these regions showed high level amplification in at least two of the analyzed cell lines. Table 4.1 summarizes the samples harbouring DNA amplification. Remarkably, 19 of the 30 cell lines studied showed DNA amplification in Region 2.

4.3.2. Integration of gene dosage and expression data

While combining frequency of gain and the presence of high level amplification events is an established approach for identifying key gene alterations in lung cancer (Tonon et al. 2005), these genes are more likely to be involved in cancer biology if they can also be demonstrated to have elevated expression levels. This is intuitive, given the accepted idea that amplified loci offer a selective advantage by conferring such over-expression (Albertson 2006). Expression levels for genes residing within the five regions displaying high level amplification defined above were determined from Affymetrix microarray data for the same NSCLC cell lines analyzed for genomic alterations. Specifically, since we aimed to identify genes with expression driven by increased gene dosage, the expression level for each of these genes (which were amplified in two or more lines) was then compared between cell lines with amplification and those that displayed neutral copy number by using the Mann-Whitney U-test. Of the 101 genes evaluated, 28 (27.7%) showed concurrent amplification and over-expression (Table 4.1). (Over-expressed genes without concurrent copy number changes were likely activated by processes other than copy number alteration.) At least eight of these 28 genes have also been described as having increased copy number in other cancer types. Recently, Yang et al. showed significant correlation between gene copy change and mRNA expression for *GNB2*, *COPS6*, and *CCT6A* on chromosome 7 in gastric cancer (Yang

2007). Additionally, concurrent copy number aberration and dysregulation of expression for *MCM7*, *NUDT1*, *CCT6A*, and *GNB2* was noted in transformed follicle centre lymphoma (Martinez-Climent et al. 2003). *EGFR* amplification has been correlated with over-expression in a variety of cancer types, including NSCLC (Reis-Filho et al. 2006; Reissmann et al. 1999; Rossi et al. 2005).

4.3.3. Quantitative RT-PCR validation in cell lines

To validate the microarray data for gene expression changes within amplicons, real-time quantitative PCR was performed on select genes shown to be amplified and over-expression in the six frequently altered regions on chromosome 7 to validate the importance of these novel regions. Gene-specific probes for *NUDT1* (Region 1), *EGFR* (Region 2), and *TAF6* (Region 5*)* were used to assess expression levels in cell lines with both gene amplification and neutral copy number in comparison to pooled normal lung cDNA. On average, *EGFR, NUDT1*, and *TAF6* gene expression was >9, >32, and >5-fold over-expresed in cell lines with amplification when compared to normal lung cDNA (respectively) and only >4, >2, and >3-fold over-expressed in cell lines with neutral copy number status compared to normal lung cDNA (respectively). Analysis by Mann-Whitney U-test showed that expression of each of these genes was significantly different between cell lines with amplification and lines with neutral copy number ($p \leq 0.05$). Figure 4.1 shows representative genome plots for cell lines with either amplification or neutral copy number status for *NUDT1* and *EGFR* (corresponding expression data for each line is also shown). These data confirmed that several distinct regions contain genes activated in NSCLC.

4.3.4. Validation of genes of interest using gene expression data for clinical NSCLC tumours

To determine if the genes identified in our study using cell lines are in fact disrupted in clinical samples, we compiled publically available gene expression microarray data for 111 clinical NSCLC samples (GEO Accession number GSE3141) (Figure 4.2). Nine genes from within these genomic loci exhibited ≥2 fold expression compared to normal bronchial epithelial (NHBE) cells in at least 45% of lung tumours (~50 cases). These

included *FTSJ2* and *NUDT1* (Region 1), *TNS3*, and *ECOP* (Region 2), and *ZNF3, TAF6, TSC22D4, MOSPD*, and *POLR2J* (Region 5).

4.3.5. Quantitative PCR validation of FTSJ2, NUDT1, TAF6, and POLR2J in clinical samples

Those genes previously implicated in cancer development and progression – *FTSJ2, NUDT1, TAF6*, and *POLR2J* – were selected for additional analysis. The *FtsJ homolog 2* (*FTSJ2)* encodes a putative RNA methyltransferase previously implicated in cell proliferation and seen to be over-expressed in lung cancer cells (Ching et al. 2002). (Other RNA methyltransferases have recently been implicated in a variety of cancer types, supported the idea that such genes are key in carcinogenesis (Frye & Watt 2006).) *Nudix (nucleoside diphosphate linked moiety X)-type motif 1* (*NUDT1)* encodes an enzyme involved in maintaining genomic integrity and has previously been demonstrated to have elevated activity in NSCLC (Speina et al. 2005; Chong et al. 2006) and several other cancer types (Kennedy et al. 1998; Okamoto et al. 1996; Wani et al. 1998; Hibi et al. 1998). Our finding is the first report of amplification-driven *NUDT1* over-expression. This is evident in both NSCLC cell lines and clinical samples. *POLR2J* (*RNA Polymerase II [DNA-directed] polypeptide J*), located in a different region of chromosome 7 observed to harbor recurring alterations, encodes a subunit of RNA Polymerase II and may contribute to transcriptional dysregulation that can drive malignant phenotypes (Fanciulli et al. 2000). The nearby *TAF6* gene, similarly activated, encodes a component of the TFIID complex, which is known to initiate transcription by RNA polymerase II (Albright & Tjian 2000; Green 2000). This complex can act as a co-activator for upstream DNA binding transcription factors and recognizes the core promoter elements of a number of cancer-associated genes including *P53*, *JUN*, *CCND1*, *GADD45*, and *P21* (Green 2000; Lu & Levine 1995; Thut et al. 1995; Bell & Tora 1999). The capacity of TAF6 to regulate genes governing a variety of key cellular processes also makes it an attractive oncogene candidate.

Further validation of expression results for the *FTSJ2*, *NUDT1, TAF6*, and *POLR2J* genes was undertaken in a panel of ten fresh-frozen NSCLC and matched normal lung clinical samples. *FTSJ2* and *NUDT1* showed higher expression in all ten tumour

samples (compared to matched normal tissue), with six samples having fold changes ≥2. *TAF6* and *POLR2J* showed higher expression in nine tumour samples compared to the matched normal tissue, with two samples having at least double the expression in tumours. All genes were significantly over-expressed relative to normal by a one-tailed Wilcoxon sign-rank test ($p \leq 0.01$) (Figure 4.3). This clear activation of putative oncogenes other than *EGFR* and *MET* supports the conclusion that lung tumourigenesis is driven by several distinct genome loci on chromosome 7. The fact that this was made clear by integration of genomic and gene expression data demonstrates the utility of moving beyond uni-dimensional analysis to identify key genes driving cancer processes.

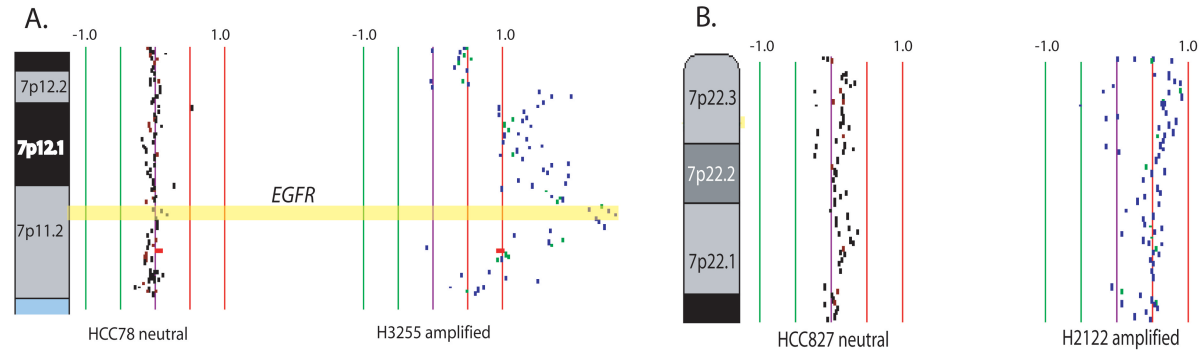
Table 4.1 – Amplified and overexpressed genes within regions of recurrent genomic gain on chromosome 7 in NSCLC cell lines.[12]

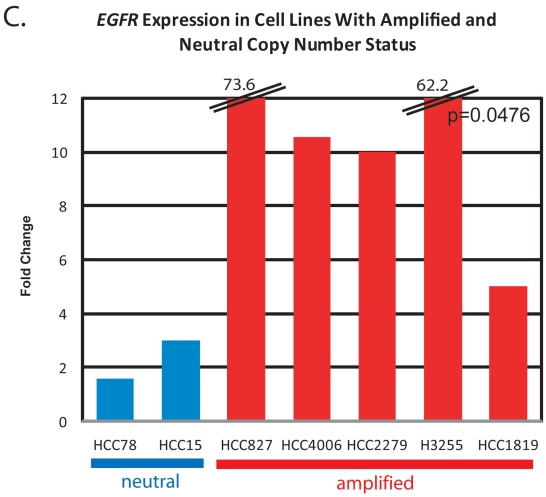
Region[2]	Base pair position	Chromosome band	Size Mbp	# of genes	Cell lines with amplification	Genes amplified and overexpressed
1	telomere - 6,162,500	7p22.3 - 7p22.1	~6	40	HCC2279, HCC193, HCC1833, HCC1195, H2122	FTSJ2, NUDT1
2	16,296,094 - 57,693,752	7p15.3 - 7p11.2	~41	164	HCC461, HCC4006, H2087, H2009, H358, H2122, H1229, H1650, H3255, HCC827, HCC2279, HCC193, H1993, HCC78, H1819, H157, HCC95	YKT6, TNS3, FIGNL1, EGFR, LANCL2, ECOP, FKBP9, MRPS17, GBAS, PSPH, CCT6A, SUMF2, CHCHD2
3	60,748,436 - 66,207,032	7q11.1 - 7q11.21	~5.4	18	none	N/A
4	71745312 - 75,278,128	7q11.23	~3.5	41	HCC193, HCC1195, H358, H1819, H1229	DKFZP434A0131
5	97,643,752 - 101,681,248	7q22.1	~4	75	HCC95, HCC1195, H2122, H1993	G10, PTCD1, ZNF38, ZNF3, COPS6, MCM7, TAF6, TSC22D4, MOSPD3, GNB2, AP1S1, POLR2J
6	156,267,184 - 158,524,992	7q36.3	~2.2	6	HCC336, HCC193	none

[12] Regions according to Garnis et al (2006).

Figure 4.1 – Representative genomic alterations within chromosome 7. A. and B. show aCGH profiles. Normalized $\log_2$ signal intensity ratios were plotted using *SeeGH* software. A $\log_2$ ratio of 0 represents equivalent copy number between the sample and reference DNA. Vertical lines denote $\log_2$ ratios from +1, +0.5, 0, -0.5, and 1, with copy number increases on the left and copy number decreases on the right of the centre line. Each black dot represents a single BAC clone. A. aCGH profile for two cell lines, one with neutral and amplified copy number status for the *EGFR* locus. B. aCGH profile for two cell lines, one with neutral and amplified copy number status for the *NUDT1* locus. qPCR results are seen in C. and D. Fold-change expression levels were calculated for cell lines with amplified (in red) and neutral copy (in blue) number status of the *EGFR* locus C. and *NUDT1* locus D. compared to a pooled normal lung cDNA reference sample. All samples were *18S*-normalized and fold change calculations were performed using the $2^{-\Delta ddCt}$ method. Significant Mann-Whitney U test *p*-values (≤ 0.05) indicate that the expression of *EGFR* and *NUDTI* is significantly different in cell lines with amplification compared to those with neutral copy number. Fold-change values are listed on the vertical axis and sample names are listed on the horizontal axis.

Figure 4.1


A.
-1.0
1.0
-1.0
1.0
7p12.2
7p12.1
7p11.2
EGFR
HCC78 neutral
H3255 amplified
B.
-1.0
1.0
-1.0
1.0
7p22.3
7p22.2
7p22.1
HCC827 neutral
H2122 amplified


C.
EGFR Expression in Cell Lines With Amplified and Neutral Copy Number Status
Fold Change
73.6
62.2
p=0.0476
HCC78
HCC15
HCC827
HCC4006
HCC2279
H3255
HCC1819
neutral
amplified

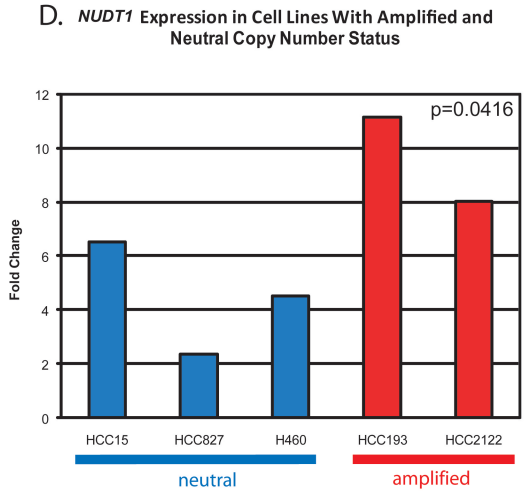

D.
NUDT1 Expression in Cell Lines With Amplified and Neutral Copy Number Status
Fold Change
p=0.0416
HCC15
HCC827
H460
HCC193
HCC2122
neutral
amplified

Figure 4.2

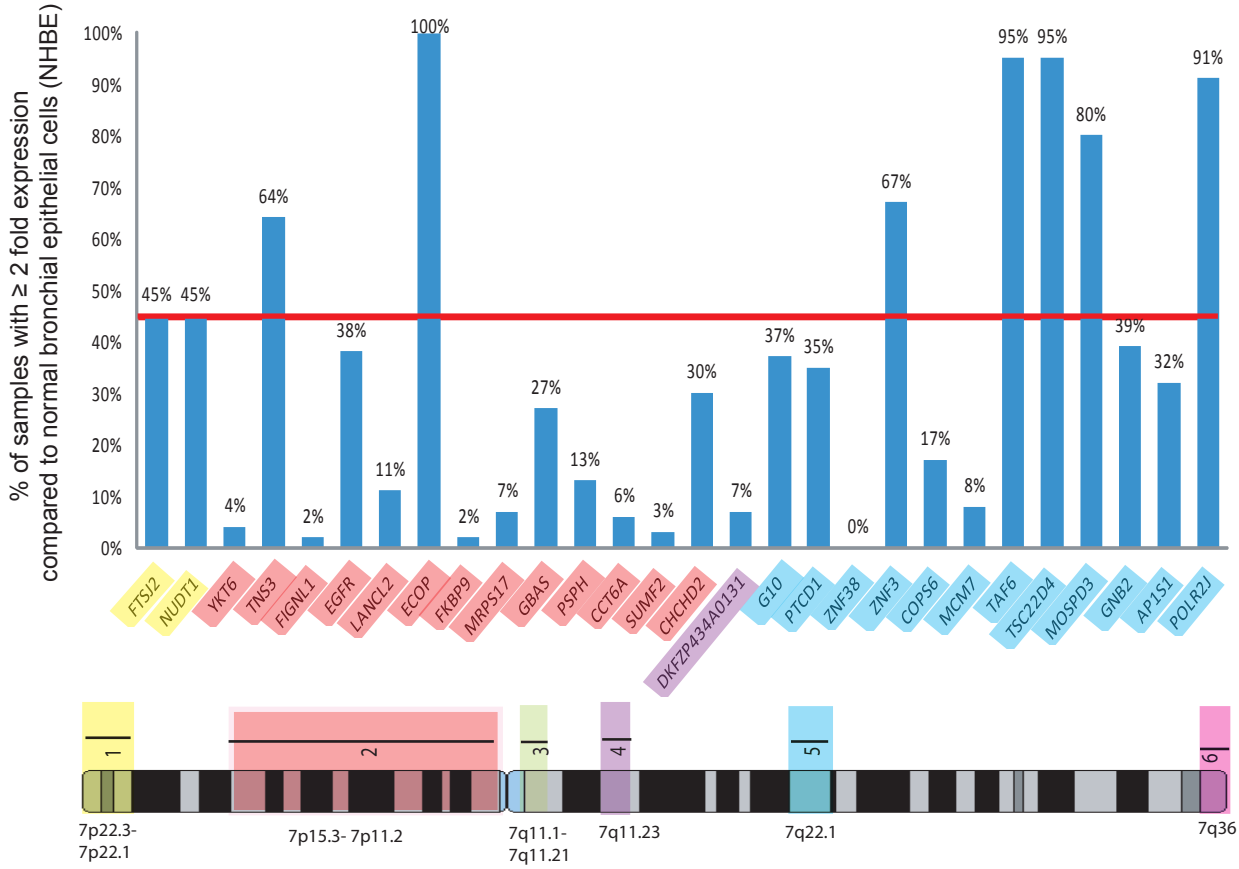


Figure 4.2 – Validation of amplified and over-expressed genes from candidate regions in a separate cohort of 111 NSCLC clinical tumours. The expression for the 28 genes from the six candidate regions which were amplified and over-expressed in the 30 NSCLC cell lines (Table 4.1) was assessed in a separate panel of 111 clinical NSCLC tumours to determine their relative expression compared to normal bronchial epithelial (NHBE) cells. Each candidate region is labelled according to its position on chromosome 7 and identified by an individual color. The respective genes from these regions which are amplified and over-expressed are presented as a histogram and color coded to match the respective region in which they are located. The percent of samples (n=111) in which each gene is ≥ 2 fold over-expressed compared to NHBE cells is presented. The red line depicts the cut-off of 45% of samples which was determined to be significant.

Figure 4.3

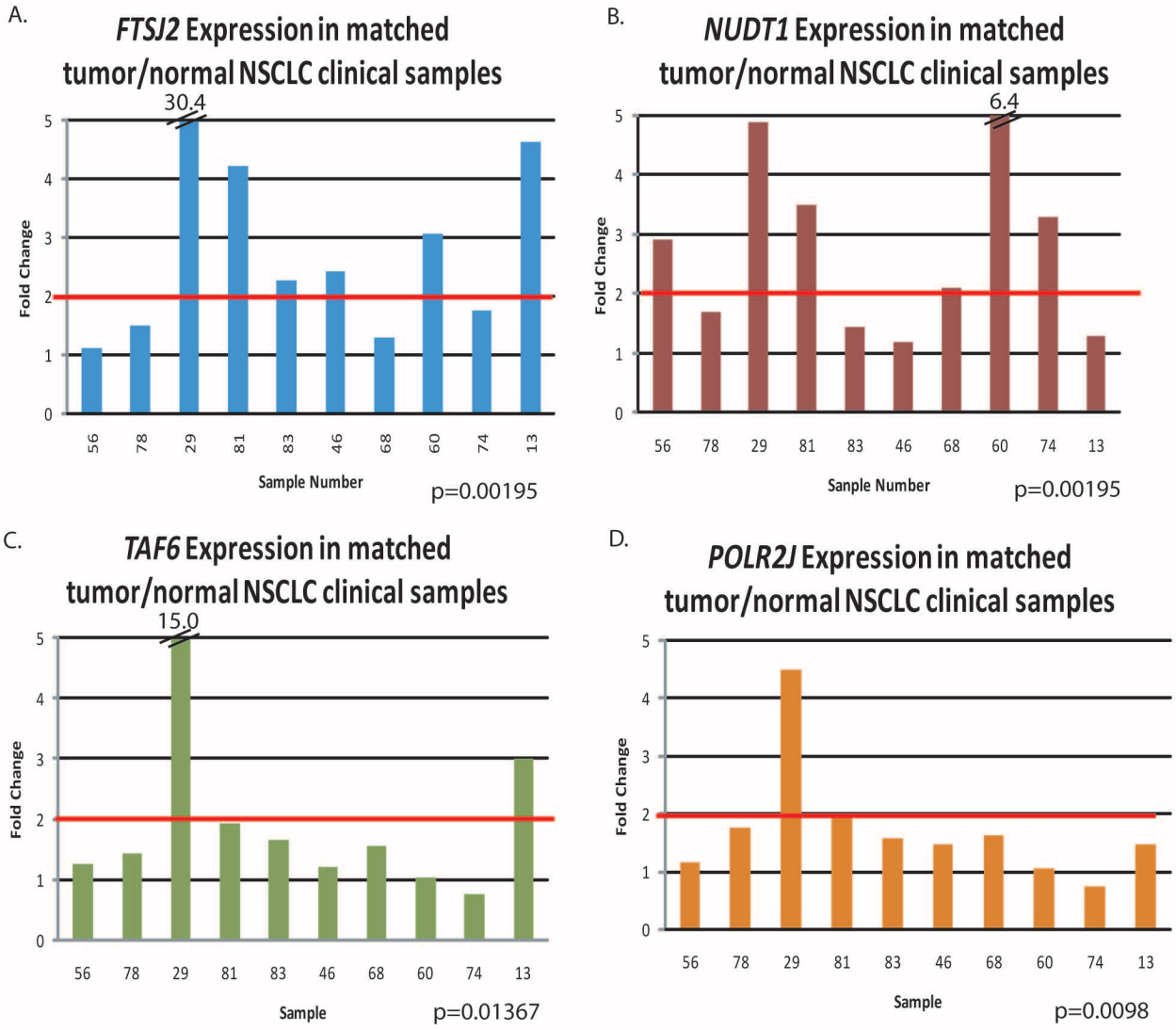


Figure 4.3 – RT-qPCR results of candidate oncogenes in ten matched tumour/normal clinical samples. *18S*-normalized cycle threshold values for the matched normal and tumour sample were used to calculate fold change values using the $2^{-\Delta ddCt}$ method. The results of a one-tailed Wilcoxon sign-rank test show that expression of *FTSJ2*, *NUDT1*, *TAF6*, and *POLR2J* is significantly different between matched tumour and normal samples ($p < 0.05$). Fold-change values are listed on the vertical axis. Fold-change values >5 are shown on the top of the graph above a hatch mark. Sample numbers are listed on the horizontal axis.

4.4. References

Albertson, D. G., 2006. Gene amplification in cancer. *Trends Genet,* 22(8), 447-55.

Albright, S. R. & R. Tjian, 2000. TAFs revisited: more data reveal new twists and confirm old ideas. *Gene,* 242(1-2), 1-13.

Arslantas, A., S. Artan, U. Oner, M. H. Muslumanoglu, M. Ozdemir, R. Durmaz, D. Arslantas, M. Vural, E. Cosan & M. A. Atasoy, 2007. Genomic alterations in low-grade, anaplastic astrocytomas and glioblastomas. *Pathol Oncol Res,* 13(1), 39-46.

Balsara, B. R. & J. R. Testa, 2002. Chromosomal imbalances in human lung cancer. *Oncogene,* 21(45), 6877-83.

Bell, B. & L. Tora, 1999. Regulation of gene expression by multiple forms of TFIID and other novel TAFII-containing complexes. *Exp Cell Res,* 246(1), 11-9.

Bild, A. H., G. Yao, J. T. Chang, Q. Wang, A. Potti, D. Chasse, M. B. Joshi, D. Harpole, J. M. Lancaster, A. Berchuck, J. A. Olson, Jr., J. R. Marks, H. K. Dressman, M. West & J. R. Nevins, 2006. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature,* 439(7074), 353-7.

Chari, R., W. W. Lockwood, B. P. Coe, A. Chu, D. Macey, A. Thomson, J. J. Davies, C. MacAulay & W. L. Lam, 2006. SIGMA: a system for integrative genomic microarray analysis of cancer genomes. *BMC Genomics,* 7, 324.

Chi, B., R. J. DeLeeuw, B. P. Coe, C. MacAulay & W. L. Lam, 2004. SeeGH--a software tool for visualization of whole genome array comparative genomic hybridization data. *BMC Bioinformatics,* 5, 13.

Chi B, D. R., Coe BP, Ng RT, MacAulay C, Lam WL, 2008. MD-SeeGH: a platform for integrative analysis of multi-dimensional genomic data. *BMC Bioinformatics,* In press.

Ching, Y. P., H. J. Zhou, J. G. Yuan, B. Q. Qiang, H. F. Kung Hf & D. Y. Jin, 2002. Identification and characterization of FTSJ2, a novel human nucleolar protein homologous to bacterial ribosomal RNA methyltransferase. *Genomics,* 79(1), 2-6.

Choi, J. S., L. T. Zheng, E. Ha, Y. J. Lim, Y. H. Kim, Y. P. Wang & Y. Lim, 2006. Comparative genomic hybridization array analysis and real-time PCR reveals genomic copy number alteration for lung adenocarcinomas. *Lung,* 184(6), 355-62.

Choi, Y. W., J. S. Choi, L. T. Zheng, Y. J. Lim, H. K. Yoon, Y. H. Kim, Y. P. Wang & Y. Lim, 2007. Comparative genomic hybridization array analysis and real time PCR reveals genomic alterations in squamous cell carcinomas of the lung. *Lung Cancer,* 55(1), 43-51.

Chong, I. W., M. Y. Chang, H. C. Chang, Y. P. Yu, C. C. Sheu, J. R. Tsai, J. Y. Hung, S. H. Chou, M. S. Tsai, J. J. Hwang & S. R. Lin, 2006. Great potential of a panel of multiple hMTH1, SPD, ITGA11 and COL11A1 markers for diagnosis of patients with non-small cell lung cancer. *Oncol Rep,* 16(5), 981-8.

Coe, B. P., W. W. Lockwood, L. Girard, R. Chari, C. Macaulay, S. Lam, A. F. Gazdar, J. D. Minna & W. L. Lam, 2006. Differential disruption of cell cycle pathways in small cell and non-small cell lung cancer. *Br J Cancer,* 94(12), 1927-35.

Engelman, J. A., K. Zejnullahu, T. Mitsudomi, Y. Song, C. Hyland, J. O. Park, N. Lindeman, C. M. Gale, X. Zhao, J. Christensen, T. Kosaka, A. J. Holmes, A. M. Rogers, F. Cappuzzo, T. Mok, C. Lee, B. E. Johnson, L. C. Cantley & P. A. Janne, 2007. MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling. *Science,* 316(5827), 1039-43.

Fanciulli, M., T. Bruno, M. Di Padova, R. De Angelis, S. Iezzi, C. Iacobini, A. Floridi & C. Passananti, 2000. Identification of a novel partner of RNA polymerase II subunit 11, Che-1, which interacts with and affects the growth suppression function of Rb. *Faseb J,* 14(7), 904-12.

Frye, M. & F. M. Watt, 2006. The RNA methyltransferase Misu (NSun2) mediates Myc-induced proliferation and is upregulated in tumors. *Curr Biol,* 16(10), 971-81.

Garcia, J. L., J. M. Hernandez, N. C. Gutierrez, T. Flores, D. Gonzalez, M. J. Calasanz, J. A. Martinez-Climent, M. A. Piris, C. Lopez-Capitan, M. B. Gonzalez, M. D. Odero & J. F. San Miguel, 2003. Abnormalities on 1q and 7q are associated with poor outcome in sporadic Burkitt's lymphoma. A cytogenetic and comparative genomic hybridization study. *Leukemia,* 17(10), 2016-24.

Garnis, C., W. W. Lockwood, E. Vucic, Y. Ge, L. Girard, J. D. Minna, A. F. Gazdar, S. Lam, C. MacAulay & W. L. Lam, 2006. High resolution analysis of non-small cell lung cancer cell lines by whole genome tiling path array CGH. *Int J Cancer,* 118(6), 1556-64.

Green, M. R., 2000. TBP-associated factors (TAFIIs): multiple, selective transcriptional mediators in common complexes. *Trends Biochem Sci,* 25(2), 59-63.

Hibi, K., Q. Liu, G. A. Beaudry, S. L. Madden, W. H. Westra, S. L. Wehage, S. C. Yang, R. F. Heitmiller, A. H. Bertelsen, D. Sidransky & J. Jen, 1998. Serial analysis of gene expression in non-small cell lung cancer. *Cancer Res,* 58(24), 5690-4.

Ishkanian, A. S., C. A. Malloff, S. K. Watson, R. J. DeLeeuw, B. Chi, B. P. Coe, A. Snijders, D. G. Albertson, D. Pinkel, M. A. Marra, V. Ling, C. MacAulay & W. L. Lam, 2004. A tiling resolution DNA microarray with complete coverage of the human genome. *Nat Genet,* 36(3), 299-303.

Jong, K., E. Marchiori, G. Meijer, A. V. Vaart & B. Ylstra, 2004. Breakpoint identification and smoothing of array comparative genomic hybridization data. *Bioinformatics,* 20(18), 3636-7.

Kennedy, C. H., R. Cueto, S. A. Belinsky, J. F. Lechner & W. A. Pryor, 1998. Overexpression of hMTH1 mRNA: a molecular marker of oxidative stress in lung cancer cells. *FEBS Lett,* 429(1), 17-20.

Khojasteh, M., W. L. Lam, R. K. Ward & C. MacAulay, 2005. A stepwise framework for the normalization of array CGH data. *BMC Bioinformatics,* 6, 274.

Lockwood, W. W., R. Chari, B. P. Coe, L. Girard, C. Macaulay, S. Lam, A. F. Gazdar, J. D. Minna & W. L. Lam, 2008. DNA amplification is a ubiquitous mechanism of oncogene activation in lung and other cancers. *Oncogene*.

Lockwood, W. W., B. P. Coe, A. C. Williams, C. MacAulay & W. L. Lam, 2007. Whole genome tiling path array CGH analysis of segmental copy number alterations in cervical cancer cell lines. *Int J Cancer,* 120(2), 436-43.

Lu, H. & A. J. Levine, 1995. Human TAFII31 protein is a transcriptional coactivator of the p53 protein. *Proc Natl Acad Sci U S A,* 92(11), 5154-8.

Martinez-Climent, J. A., A. A. Alizadeh, R. Segraves, D. Blesa, F. Rubio-Moscardo, D. G. Albertson, J. Garcia-Conde, M. J. Dyer, R. Levy, D. Pinkel & I. S. Lossos, 2003. Transformation of follicular lymphoma to diffuse large cell lymphoma is associated with a heterogeneous set of DNA copy number and gene expression alterations. *Blood,* 101(8), 3109-17.

Okamoto, K., S. Toyokuni, W. J. Kim, O. Ogawa, Y. Kakehi, S. Arao, H. Hiai & O. Yoshida, 1996. Overexpression of human mutT homologue gene messenger RNA in renal-cell carcinoma: evidence of persistent oxidative stress in cancer. *Int J Cancer,* 65(4), 437-41.

Panani, A. D. & C. Roussos, 2006. Cytogenetic and molecular aspects of lung cancer. *Cancer Lett,* 239(1), 1-9.

Pei, J., B. R. Balsara, W. Li, S. Litwin, E. Gabrielson, M. Feder, J. Jen & J. R. Testa, 2001. Genomic imbalances in human lung adenocarcinomas and squamous cell carcinomas. *Genes Chromosomes Cancer,* 31(3), 282-7.

Reis-Filho, J. S., C. Pinheiro, M. B. Lambros, F. Milanezi, S. Carvalho, K. Savage, P. T. Simpson, C. Jones, S. Swift, A. Mackay, R. M. Reis, J. L. Hornick, E. M. Pereira, F. Baltazar, C. D. Fletcher, A. Ashworth, S. R. Lakhani & F. C. Schmitt, 2006. EGFR amplification and lack of activating mutations in metaplastic breast carcinomas. *J Pathol,* 209(4), 445-53.

Reissmann, P. T., H. Koga, R. A. Figlin, E. C. Holmes & D. J. Slamon, 1999. Amplification and overexpression of the cyclin D1 and epidermal growth factor receptor genes in non-small-cell lung cancer. Lung Cancer Study Group. *J Cancer Res Clin Oncol,* 125(2), 61-70.

Rossi, M. R., J. La Duca, S. Matsui, N. J. Nowak, L. Hawthorn & J. K. Cowell, 2005. Novel amplicons on the short arm of chromosome 7 identified using high resolution array CGH contain over expressed genes in addition to EGFR in glioblastoma multiforme. *Genes Chromosomes Cancer,* 44(4), 392-404.

Sequist, L. V., D. W. Bell, T. J. Lynch & D. A. Haber, 2007. Molecular predictors of response to epidermal growth factor receptor antagonists in non-small-cell lung cancer. *J Clin Oncol,* 25(5), 587-95.

Shadeo, A. & W. L. Lam, 2006. Comprehensive copy number profiles of breast cancer cell model genomes. *Breast Cancer Res,* 8(1), R9.

Speina, E., K. D. Arczewska, D. Gackowski, M. Zielinska, A. Siomek, J. Kowalewski, R. Olinski, B. Tudek & J. T. Kusmierek, 2005. Contribution of hMTH1 to the maintenance of 8-oxoguanine levels in lung DNA of non-small-cell lung cancer patients. *J Natl Cancer Inst,* 97(5), 384-95.

Thut, C. J., J. L. Chen, R. Klemm & R. Tjian, 1995. p53 transcriptional activation mediated by coactivators TAFII40 and TAFII60. *Science,* 267(5194), 100-4.

Tonon, G., K. K. Wong, G. Maulik, C. Brennan, B. Feng, Y. Zhang, D. B. Khatry, A. Protopopov, M. J. You, A. J. Aguirre, E. S. Martin, Z. Yang, H. Ji, L. Chin & R. A. Depinho, 2005. High-resolution genomic profiles of human lung cancer. *Proc Natl Acad Sci U S A,* 102(27), 9625-30.

Travis, W. D., T. V. Colby, B. Corrin, Y. Shimosato & E. Brambilla, 1999. *Histological Typing of Lung and Pleural Tumours with contributions by Pathologists from 14 Countries,* Berlin: Springer Verlag.

Waldman, F. M., P. R. Carroll, R. Kerschmann, M. B. Cohen, F. G. Field & B. H. Mayall, 1991. Centromeric copy number of chromosome 7 is strongly correlated with tumor grade and labeling index in human bladder cancer. *Cancer Res,* 51(14), 3807-13.

Wani, G., G. E. Milo & S. M. D'Ambrosio, 1998. Enhanced expression of the 8-oxo-7,8-dihydrodeoxyguanosine triphosphatase gene in human breast tumor cells. *Cancer Lett,* 125(1-2), 123-30.

Watson, S. K., R. J. deLeeuw, D. E. Horsman, J. A. Squire & W. L. Lam, 2007. Cytogenetically balanced translocations are associated with focal copy number alterations. *Hum Genet,* 120(6), 795-805.

Wilson, I. M., J. J. Davies, M. Weber, C. J. Brown, C. E. Alvarez, C. MacAulay, D. Schubeler & W. L. Lam, 2006. Epigenomics: mapping the methylome. *Cell Cycle,* 5(2), 155-8.

Yang, S., 2007. Gene amplifications at chromosome 7 of the human gastric cancer genome. *Int J Mol Med,* 20(2), 225-31.

Zhou, B. B., M. Peyton, B. He, C. Liu, L. Girard, E. Caudler, Y. Lo, F. Baribaud, I. Mikami, N. Reguart, G. Yang, Y. Li, W. Yao, K. Vaddi, A. F. Gazdar, S. M. Friedman, D. M. Jablons, R. C. Newton, J. S. Fridman, J. D. Minna & P. A. Scherle, 2006. Targeting ADAM-mediated ligand cleavage to inhibit HER3 and EGFR pathways in non-small cell lung cancer. *Cancer Cell,* 10(1), 39-50.

Zojer, N., G. Dekan, J. Ackermann, M. Fiegl, H. Kaufmann, J. Drach & H. Huber, 2000. Aneuploidy of chromosome 7 can be detected in invasive lung cancer and associated premalignant lesions of the lung by fluorescence in situ hybridisation. *Lung Cancer,* 28(3), 225-35.

5. DEFINING GENOMIC ALTERATION BOUNDARIES FOR A COMBINED SMALL CELL AND NON-SMALL CELL LUNG CARCINOMA[13]

5.1. Introduction

Because pulmonary metastases and multiple independent primary tumors will be staged and managed differently, determining the clonal relationship between multiple tumors found in the same patient is essential. In clinical practice, the relationship between such tumors may be defined by their location and morphologic features (Martini & Melamed 1975). Unfortunately, histopathological criteria may not be adequate to determine clonality with sufficient certainty. Tumors that exhibit similar histology may not easily be defined as independent or related. Moreover, it is possible that different cancer subtypes arise from the same progenitor cell, suggesting that it may not be appropriate to use a default diagnosis of multiple primary tumors where subtypes differ (Petersen & Petersen 2001).

A small cell lung carcinoma (SCLC) that contains histologically distinct areas of one or more types of non-small cell carcinoma is called a combined tumor. The non-small cell component could be adenocarcinoma (AC), squamous carcinoma (SQ), large cell carcinoma (LC) or large cell neuroendocrine carcinoma (LCNEC). Although combined tumors are rare, multiple synchronous and metachronous lung cancers are encountered more frequently, especially with the current use of highly sensitive spiral computed tomography (CT) for early detection of lung cancer. Clonal relationships between components of combined tumors involving SCLC have been investigated using techniques that assess allelic genomic imbalances and/or oncogenic mutations commonly found in lung cancer (Huang et al. 2002; Murase et al. 2003; Fellegara et al. 2008; D'Adda et al. 2008). However, these techniques can only assess imbalanced losses at a limited number of chromosomal loci, providing very low resolution of genomic alteration that occurs within the tumor cell genome.

[13] A version of this chapter has been published: Buys TPH, Aviel-Ronen S, Waddell TK, Lam WL, and Tsao MS. (**2009**) "Defining genomic alteration boundaries for a combined small cell and non-small cell lung carcinoma" *Journal of Thoracic Oncology*, 4(2):227-39.

Characterizing the boundaries of genomic alteration for these combined tumors may provide some insight into the origin and clonal evolution of multiple histologically distinct tumors. Genomic alterations shared among different histological components may represent features that characterize the cancer "stem/progenitor" cells of the tumor. These genomic changes could possibly account for shared tumor phenotype, such as resistance and may represent what should be the true targets of therapy.

The advent of high-resolution genome profiling technologies represents a powerful tool for defining the clonal relationship between tumors arising in a single patient. Recent reports have demonstrated the utility of array comparative genomic hybridization (CGH) in establishing the relationship between tumors in such cases, including a survey of multiple lung tumors from three different patients (Gallegos Ruiz et al. 2007) In this report, we have applied a similar strategy to defining the relationship between a combined lung cancer with at least three histological components: SCLC, AC, and LCNEC. In addition, the SCLC demonstrated two components that are distinguishable by their divergent immunohistochemistry phenotypes. Based on a tiling-path platform comprised of overlapping elements that span the human genome, we have fine mapped boundaries for segmental DNA alterations in each of these component tumors, using the presence of identical genomic breakpoints between tumors as signature markers for shared clonal origin.

5.1.1. Case report

A 71-year old man developed pain in his neck, leading to chest x-ray and CT studies that showed a right upper lobe lung mass. A CT-guided fine needle aspiration biopsy showed non-small cell lung cancer. He smoked 1–1.5 packs per day for 27 years until quitting 14 years before presentation. Extra-thoracic staging investigation with abdominal ultrasound showed multiple echogenic nodules in the liver, but similar findings were noted 18 months earlier and CT scan evaluation of one of these lesions was felt to be a hemangioma. The patient was also noted to have a right adrenal myelolipoma both 18 months earlier and at staging investigation. Although a bone scan showed a small focus of activity in the mid-cervical spine on the left side, it was felt to

be non-specific and head CT was normal. Following negative frozen section evaluation of mediastinal lymph node biopsies, a right upper lobe lobectomy was performed. The patient's post-operative course was complicated by pneumonia that required outpatient antibiotic treatment. Based on the pathology report, post-operative adjuvant chemotherapy was recommended, however prior to initiating treatment, re-staging examinations demonstrated metastatic spread to the sternum and to the liver. A destructive bone lesion was identified in the mid-manubrium, which clearly had not been present prior to surgery just a few weeks earlier. Similarly, the appearance of the liver changed dramatically, as he developed liver function test abnormalities. Chemotherapy with carboplatin and etoposide resulted in some symptomatic response but disease progression of the metastases by imaging. The metastatic deposits have not been biopsied to evaluate which of the tumor components has metastasized.

5.2. Materials and methods

5.2.1. Immunohistochemistry

Immunohistochemistry studies were performed using the Benchmark automated immunohistochemistry equipment (Ventana Medical System, Tucson, AZ). Pre-diluted antibodies (Ventana) for TTF-1, chromogranin-A, synaptophysin, and p53 were used. Cytokeratin 7 (Dako Canada, Mississauga, ON), CD56 (Zymed Laboratories, South San Francisco, CA) and p63 (Vector Laboratories, Burlingame, CA) were used at 1:2000, 1:100, and 1:50 dilutions. Microwave heat pre-treatment was performed for TTF-1, CD56, synaptophysin, p53, and p63, while protease and no pre-treatment were done for cytokeratin 7 and chromogranin-A, respectively.

5.2.2. DNA extraction

Tumor and matched normal (non-cancerous) lung tissue samples were obtained at the resection specimen. Individual tumor samples were cored from specific areas identified from corresponding sections stained by hematoxylin-eosin and immunohistochemistry, using the 1.5 mm coring needles of Tissue Arrayer (Beecher Instruments, Sun Prairie, WI). Tissue cores were obtained for DNA extraction following histopathological review

of a hematoxylin-and-eosin stained reference slide. DNA was extracted by a standard phenol:chloroform protocol. Briefly, deparaffinised tissue cores were digested overnight with Proteinase K in a buffer containing 10 mM Tris-HCl (pH 8), 1 mM EDTA, 0.5% SDS, and 50 mM NaCl. Subsequent treatment with RNase A was followed by extraction with a phenol:chloroform mixture, ethanol precipitation, and re-suspension in H_2O. DNA concentrations were determined using an ND-1000 Spectrophotometer (Thermo Scientific, Wilmington, DE).

5.2.3. Whole genome tiling-path array CGH

The SMRTr array, comprised of 26,363 overlapping elements, was manufactured at the British Columbia Cancer Agency, as previously described (Watson et al. 2007) Separate genome profiles were generated for each tumor subcomponent, as well as for matched normal tissue. Briefly, 400 ng of patient DNA and 400 ng of individual male genomic reference DNA were labelled with cyanine-3 and cyanine-5 dyes respectively (Garnis et al. 2003). DNA probes were then pooled and unincorporated nucleotides were removed with a YM-30 Microcon centrifugation tube (Millipore, Etobicoke, ON). One hundred µg of Cot-1 DNA (Invitrogen, Carlsbad, CA) was then added and the entire mixture was precipitated. This material was then re-suspended in a 45 µl cocktail consisting of DIG Easy hybridization solution (Roche, Mississaga, ON), sheared herring sperm DNA (Sigma-Aldrich, Oakville, ON), and yeast tRNA (Calbiochem, Mississaga, ON). Probe denaturing and blocking steps were then carried out at 85°C and 45°C for 10 minutes and for one hour respectively. The probe mixture was then applied to the surface of the SMRTr arrays, coverslips were affixed, and arrays were incubated at 45°C for 36-40 hours. Post-incubation, slides underwent five agitating washes in 0.1X saline sodium citrate, 0.1% SDS at 45°C (each wash ~5 min). Rinses with 0.1X SSC followed, then drying by centrifugation. The patient blood DNA sample was also hybridized against the same individual male genomic reference sample.

5.2.4. Sample imaging and data analysis

Array images were obtained with a CCD camera system and analyzed with SoftWoRx Tracker Spot Analysis software (Applied Precision, Issaquah, WA). Experimental bias

due to the array platform and the use of archived tissue specimens were normalized as previously described (Khojasteh et al. 2005; Chi et al. 2008). SeeGH software was used to plot log2 signal intensity ratios for each clone on the array against chromosomal position (Chi et al. 2008). BAC array data were filtered from analysis where standard deviations between replicate spots were >0.1 or signal-to-noise ratios were <3. Consensus values from replicate analyses using the aCGH-Smooth segmentation algorithm were used to facilitate definition of genome breakpoint boundaries (Jong et al. 2004). Alterations associated with normal (non-cancerous) tissue were excluded from analysis. Array CGH data have been made publicly available through the System for Integrative Genomic Microarray Analysis (SIGMA) resource (http://sigma.bccrc.ca) (Chari et al. 2006).

5.3. Results

5.3.1. Tumor histopathology

The right upper lobe showed a 4.5 cm tumor with at least three histologically different tumor areas. The predominant component (estimated at 60%) was a large cell carcinoma with focal trabecular and insular growth pattern (Figure 5.1A), prominent necrosis (Figure 5.1B), >10/10 high power field mitotic count and high percentage areas of tumor cells showing membranous staining for CD56 (Figure 5.1C) but negative staining for chromogranin-A and synaptophysin. Despite the negative staining for the two latter markers, the histology and CD56 expression were consistent with a LCNEC. Approximately 25% of the tumor was a mixed type moderately differentiated AC with predominantly papillary and acinar patterns (Figure 5.1D). The remaining tumor showed a SC carcinoma histopathology with scant cytoplasm, hyperchromatic nuclei, and inconspicuous nucleoli (Figure 5.1E and 1F). In addition, this SC tumor showed two areas that were distinguished by only focally positive chromogranin-A (Chr)/CD56+ (SC1, Figure 5.1G) and chr+/CD56- (Figure 5.1H) immunoprofiles. The immunophenotypes of the various tumor components are detailed in Table 5.1. With mediastinal lymph node stations 2R, 4R, 4L, 7, and 11 being negative for malignancy,

the tumor was pathologically staged as pT2N0Mx with prominent lymphatic and vascular invasion.

5.3.2. Comparative genomic profiles

Several segmental DNA changes were detected for the AC, LCNEC, and each SC component (SC1, SC2) (Table 5.2, Supplementary Figure 5.1[14]). Genome alterations previously associated with each lung tumor subtype were observed, including loss within chromosome 13q that has been associated with LCNEC tumor and gains within chromosome 5p and losses encompassing 17p in SCLC (both SC1 and SC2) (Ullmann et al. 1998; Ullmann et al. 2001; Balsara & Testa 2002; Peng et al. 2005). Alterations commonly associated with AC included gains within chromosome 7p (including *EGFR*) and 17q, as well as loss within chromosome 3p (Choi et al. 2006). Also noted were alterations with different boundaries seen in different tumors, yet encompassing similar genomic loci. The 1q21.2-q41 region was gained in both the LCNEC and SCLC components, the former having a segmental gain within the chromosome arm and the other tumors harboring whole arm gains. Similarly, both the LCNEC and AC tumors appear to have acquired additional copies of the *MYC* oncogene at chromosome 8q24.21, with the former having a focal amplification in a broader segmental gain spanning this gene and the latter having a more narrowly defined segmental gain (Figure 5.2A).

Identical genomic alterations in multiple tumor types, specifically, SC1 and SC2 shared multiple whole chromosome arm changes, including gain of chromosome arms 1q and 5p, loss of chromosome arm 15q and multiple intra-chromosomal breakpoints. The latter included an identical breakpoint within chromosome band 13q21.33, with both tumors showing a loss from 13q11-q21.33 and a gain from 13q21.33-q34 (Figure 5.2B). A fragile site for this specific genomic region has not been reported previously (Durkin & Glover 2007; Lukusa & Fryns 2007). Furthermore, this alteration boundary was not observed in any tiling-path array CGH profiles for lung cancer cell lines (Coe & Lockwood et al. 2006; Garnis et al. 2006) and has not been reported even at low

[14]All supplementary data for this manuscript may be found at the journal website (http://www.jto.org).

frequency in other high resolution analyses of lung tumor genomes (Choi et al. 2006; Kim et al. 2005; Peng et al. 2005; Shibata et al. 2005; Tonon et al. 2005; Weir et al. 2007).

Molecular alterations in SC1 and SC2 were not completely identical, as differences between these tumors were observed at other genomic loci. Analysis of both individual profiles for SC1 (Chr-/CD56+) and SC2 (Chr+/CD56-) and a direct comparative hybridization between these two tumors confirmed a 20p deletion specific to the SC1 tumor (Figure 5.2C). Gene copy number changes at *CHGA* (chromogranin-A) or *NCAM1* (CD56) were not observed for either SCLC subcomponent.

5.4. Discussion

Many of the genome alterations observed in this case matched previously described changes associated with specific lung cancer histologies. Those genes altered in multiple tumors from the same individual may represent changes that are essential to lung tumorigenesis, regardless of disease subtype. Where alterations with different boundaries spanned the same genomic regions, independent activation of these regions can be inferred. This suggests that genes in these regions may be necessary factors for lung cancer initiation in general. In this case, gain at chromosome 1q21.2-q41 in the LCNEC, SC1, and SC2 components suggested independent activation of several genes, including the putative oncogene *KIF14*, which has previously been implicated in a variety of cancer types (Corson et al. 2005; Corson et al. 2007). The *MYC* oncogene appears to have undergone similar independent activation in the AC and LCNEC tumors.

The fine mapping of alteration boundaries afforded by our high-resolution genome analysis also revealed identical alterations for the SCLC components of this case. Given the precise similarity of these alteration boundaries, their lack of association with previously described genome fragile sites, and their rarity in previously undertaken genome analyses of lung cancer cells, it is quite possible that the identical chromosomal rearrangements found in these tumors represent an early evolution from a common tumor stem/progenitor cell. Changes such as deletion of chromosome 20p in SC1 but

not in SC2 would then be indicative of subsequent clonal divergence. While morphologic features are limited for evaluating the relationship between SC1 and SC2, the genomic data are salient with respect to confirming the clonal relationship between these lesions.

Taken together, our results suggest that this patient had multiple primary tumors with divergent clonal evolution. Based on alterations seen in SC1 and SC2 and the fact that the LCNEC and adenocarcinoma components harbored several alterations not observed in the other tumors, we deduced a clonal relationship between tumors that is depicted in Figure 5.2D. This case demonstrates the potential utility of using DNA alteration boundaries as a means of classifying tumor relationships. Ultimately, integration of our data with global analysis of DNA sequence changes or epigenetic alterations for these same tumors could yield even further insights into the clonal relationship between these tumors (Sjoblom 2008; Wilson et al. 2006). With standardized approaches to sample handling and data analysis, high resolution genome profiling technologies may be reasonably adapted to a clinical setting where they could have a significant impact on diagnosis and possibly on patient care.

Figure 5.1 – Microscopic features of a combined carcinoma. A. Bordering area of large cell neuroendocrine carcinoma (LCNEC) and adenocarcinoma components. B. Higher magnification of LCNEC showing focal necrosis. C. CD56+ membranous immunostaining for LCNEC tumor cells. D. Adenocarcinoma component with papillary and acinar features. E.-H. Small cell carcinoma component of combined tumor with two areas distinguished by their expression of CD56 and chromogranin-A (Chr). SC1 area showed negative or focal chromogranin-A staining (H.) and strong membranous CD56 staining (G.), while SC2 was chromogranin-A positive (H.) but CD56 negative (G.). (Magnifications: A. and E., x25; B., C., and F., x200; D., x100; A., B., C., D., F. Staining: hematoxylin and eosin; C., G., H.: immunohistochemistry).

Figure 5.1

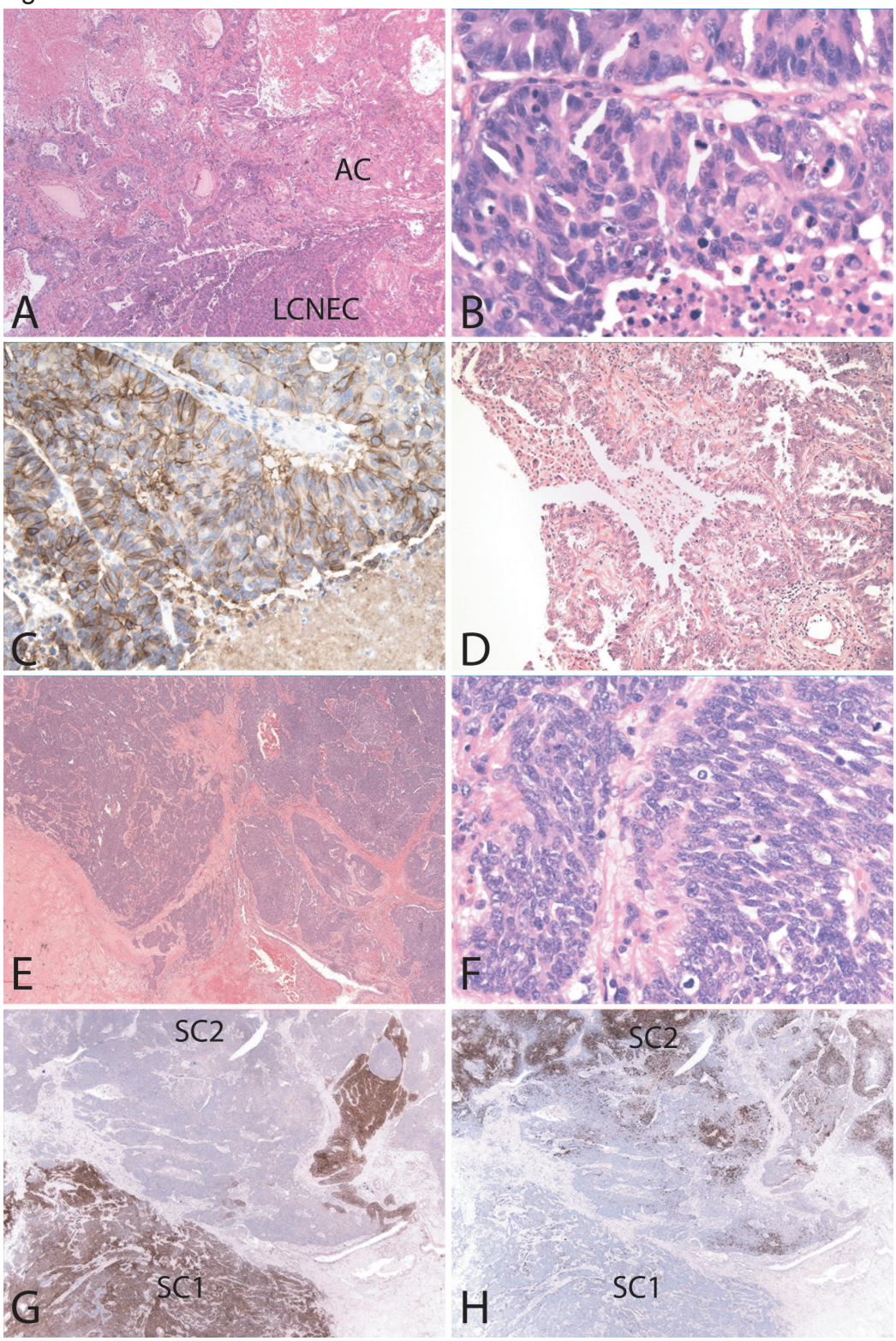


Figure 5.2 – Comparison of genomic alterations between subcomponents of a combined carcinoma. Each clone on the array CGH platform is represented at its known chromosomal position by a small dark blue line. For each array experiment, the signal intensity ratio at each clone is plotted on a log2 scale along the horizontal axis. Vertical lines show log2 signal intensity ratios from -1.0 (green line) to +1.0 (red line), with unchanged regions falling along the zero line (purple), green shading marking regions of segmental genomic loss, and red shading marking regions of segmental gain. A. Two tumors harbored additional copies of the MYC oncogene: a focal amplification spanning this gene was seen for the LCNEC component while a lower level segmental gain was seen for the AC tumor. B. An example of a chromosomal breakpoint shared by SC1 and SC2. Neither the AC nor LCNEC components exhibited this alteration boundary within 13q21.33. C. The alterations status of chromosome 20 was different between SC1 and SC2, suggesting ongoing evolution. The first two karyograms show genome profiles generated by comparison of tumor and reference DNA and are annotated as in parts A. and B. The third karyogram (within the gray box) shows the results of a direct competitive hybridization experiment between SC1 and SC2. As above, where clones are positioned along the zero line (purple), equivalent copy number between the two samples is inferred. Where log2<0, it suggests increased DNA copy number for SC2 relative to SC1 (green shading). Where log2>0, the reverse may be inferred. D. A putative clonal relationship of genetically distinct tumor types in this combined carcinoma. Tumors could have arisen from different normal airway progenitor cells that might be pleuripotent or lineage-defined.

Figure 5.2

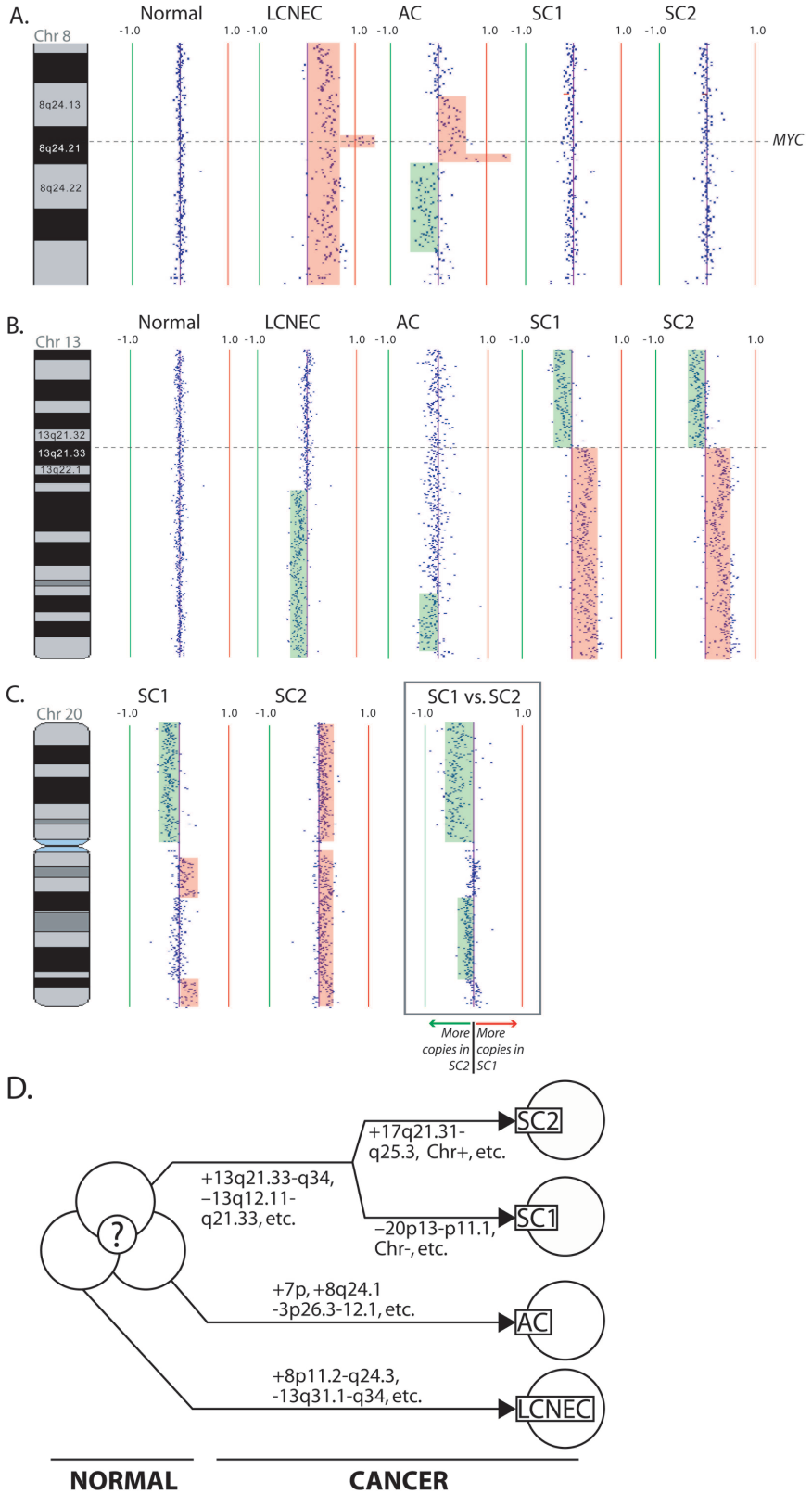

A.
Normal
LCNEC
AC
SC1
SC2
Chr 8
8q24.13
8q24.21
8q24.22
MYC
B.
Chr 13
13q21.32
13q21.33
13q22.1
C.
Chr 20
SC1 vs. SC2
More copies in SC2
More copies in SC1
D.
+17q21.31-q25.3, Chr+, etc.
+13q21.33-q34, –13q12.11-q21.33, etc.
–20p13-p11.1, Chr-, etc.
+7p, +8q24.1 -3p26.3-12.1, etc.
+8p11.2-q24.3, -13q31.1-q34, etc.
SC2
SC1
AC
LCNEC
?
NORMAL
CANCER

Table 5.1 – Immunophenotypes of tumor components studied by array CGH.

	LCNEC	AD	SC1	SC2
CK7	+	+	+	+
TTF-1	+	+	+	+
CD56	+	-	+	-
Chromogranin-A	-	-	Focal	+
Synaptophysin	-	-	-	+
P63	-	-	-	-
P53	+	-	-	-

LCNEC, large cell neuroendocrine carcinoma; AD, adenocarcinoma, SC, small cell carcinoma.

Table 5.2 – Summary of genome alterations identified for each tumor component.

Tumor component	Shared alterations	Independent alterations
AC	None identified	Gain of 4p14-p13, 5p15.33, 6p22.2-21.2, 6q22.33-q23.3, 6q27, 7p21.3-p11.2, 8q24.13-q24.21, 16p13.3-p13.2, 17q12-q25.3, 20q13.2, 20q13.33, 22q11.21-q13.33 Loss of 3p26.3-p22.1, 4p16.1-p15.1, 8p23.1, 9p24.3-q33.1, 11p15.4, 13q33.1-q34
LCNEC	None identified	Gain of 1q21.2-q41, 8p12-q24.3, 10p15.3-p11.23 Loss of 3q12.3-q21.3, 6q12-q27, 8p23.3-p12, 13q31.1-q34, 16q23.1-q23.3, 17p13.2-p12
SC1 (Chr-/ CD56+)	Gain of 1q21.1-q44*, 5p15.33-p12*, 6p25.3-p21.2, 7p22.2-p14.1, 13q21.33-q34 Loss of 13q11-q21.33, 15q11.1-q26.3*	Gain of 16p13.3-p11.2, 20q11.23-q12, 20q13.32-q13.33 Loss of 4q31.3, 20p13-p11.21, 17p13.2-p11.2*
SC2 (Chr+/ CD56-)	Gain of 1q21.1-q44*, 5p15.33-p12*, 6p25.3-p21.2, 7p22.2-p14.1, 13q21.33-q34 Loss of 13q11-q21.33, 15q11.1-q26.3*	Gain of 20p13-q13.33* Loss of 10q11.1-q26.3, 12q12, 21q11.1-q11.2, 22q11.21-q13.33, 17p13.2-q21.31

*Denotes whole chromosome arm or whole chromosome change.

5.5. References

Balsara, B. R. & J. R. Testa, 2002. Chromosomal imbalances in human lung cancer. *Oncogene,* 21(45), 6877-83.

Chari, R., W. W. Lockwood, B. P. Coe, A. Chu, D. Macey, A. Thomson, J. J. Davies, C. Macaulay & W. L. Lam, 2006. SIGMA: A System for Integrative Genomic Microarray Analysis of Cancer Genomes. *BMC Genomics,* 7(1), 324.

Chi, B., R. J. deLeeuw, B. P. Coe, R. T. Ng, C. MacAulay & W. L. Lam, 2008. MD-SeeGH: a platform for integrative analysis of multi-dimensional genomic data. *BMC Bioinformatics,* 9, 243.

Choi, J. S., L. T. Zheng, E. Ha, Y. J. Lim, Y. H. Kim, Y. P. Wang & Y. Lim, 2006. Comparative genomic hybridization array analysis and real-time PCR reveals genomic copy number alteration for lung adenocarcinomas. *Lung,* 184(6), 355-62.

Coe, B. P., W. W. Lockwood, L. Girard, R. Chari, C. Macaulay, S. Lam, A. F. Gazdar, J. D. Minna & W. L. Lam, 2006. Differential disruption of cell cycle pathways in small cell and non-small cell lung cancer. *Br J Cancer,* 94(12), 1927-35.

Corson, T. W., A. Huang, M. S. Tsao & B. L. Gallie, 2005. KIF14 is a candidate oncogene in the 1q minimal region of genomic gain in multiple cancers. *Oncogene,* 24(30), 4741-53.

Corson, T. W., C. Q. Zhu, S. K. Lau, F. A. Shepherd, M. S. Tsao & B. L. Gallie, 2007. KIF14 messenger RNA expression is independently prognostic for outcome in lung cancer. *Clin Cancer Res,* 13(11), 3229-34.

D'Adda, T., G. Pelosi, C. Lagrasta, C. Azzoni, L. Bottarelli, S. Pizzi, I. Troisi, G. Rindi & C. Bordi, 2008. Genetic alterations in combined neuroendocrine neoplasms of the lung. *Mod Pathol,* 21(4), 414-22.

Durkin, S. G. & T. W. Glover, 2007. Chromosome fragile sites. *Annu Rev Genet,* 41, 169-92.

Fellegara, G., T. D'Adda, F. P. Pilato, E. Froio, L. Ampollini, M. Rusca & G. Rindi, 2008. Genetics of a combined lung small cell carcinoma and large cell neuroendocrine carcinoma with adenocarcinoma. *Virchows Arch,* 453(1), 107-15.

Gallegos Ruiz, M. I., H. van Cruijsen, E. F. Smit, K. Grunberg, G. A. Meijer, J. A. Rodriguez, B. Ylstra & G. Giaccone, 2007. Genetic heterogeneity in patients with multiple neoplastic lung lesions: a report of three cases. *J Thorac Oncol,* 2(1), 12-21.

Garnis, C., C. Baldwin, L. Zhang, M. P. Rosin & W. L. Lam, 2003. Use of complete coverage array comparative genomic hybridization to define copy number alterations on chromosome 3p in oral squamous cell carcinomas. *Cancer Res,* 63(24), 8582-5.

Garnis, C., W. W. Lockwood, E. Vucic, Y. Ge, L. Girard, J. D. Minna, A. F. Gazdar, S. Lam, C. Macaulay & W. L. Lam, 2006. High resolution analysis of non-small cell lung cancer cell lines by whole genome tiling path array CGH. *Int J Cancer,* 118(6), 1556-64.

Huang, J., C. Behrens, Wistuba, II, A. F. Gazdar & J. Jagirdar, 2002. Clonality of combined tumors. *Arch Pathol Lab Med,* 126(4), 437-41.

Jong, K., E. Marchiori, G. Meijer, A. V. Vaart & B. Ylstra, 2004. Breakpoint identification and smoothing of array comparative genomic hybridization data. *Bioinformatics,* 20(18), 3636-7.

Khojasteh, M., W. L. Lam, R. K. Ward & C. MacAulay, 2005. A stepwise framework for the normalization of array CGH data. *BMC Bioinformatics,* 6, 274.

Kim, T. M., S. H. Yim, J. S. Lee, M. S. Kwon, J. W. Ryu, H. M. Kang, H. Fiegler, N. P. Carter & Y. J. Chung, 2005. Genome-wide screening of genomic alterations and their clinicopathologic implications in non-small cell lung cancers. *Clin Cancer Res,* 11(23), 8235-42.

Lukusa, T. & J. P. Fryns, 2007. Human chromosome fragility. *Biochim Biophys Acta*.

Martini, N. & M. R. Melamed, 1975. Multiple primary lung cancers. *J Thorac Cardiovasc Surg,* 70(4), 606-12.

Murase, T., H. Takino, S. Shimizu, H. Inagaki, H. Tateyama, E. Takahashi, H. Matsuda & T. Eimoto, 2003. Clonality analysis of different histological components in combined small cell and non-small cell carcinoma of the lung. *Hum Pathol,* 34(11), 1178-84.

Peng, W. X., T. Shibata, H. Katoh, A. Kokubu, Y. Matsuno, H. Asamura, R. Tsuchiya, Y. Kanai, F. Hosoda, T. Sakiyama, M. Ohki, I. Imoto, J. Inazawa & S. Hirohashi, 2005. Array-based comparative genomic hybridization analysis of high-grade neuroendocrine tumors of the lung. *Cancer Sci,* 96(10), 661-7.

Petersen, I. & S. Petersen, 2001. Towards a genetic-based classification of human lung cancer. *Anal Cell Pathol,* 22(3), 111-21.

Shibata, T., S. Uryu, A. Kokubu, F. Hosoda, M. Ohki, T. Sakiyama, Y. Matsuno, R. Tsuchiya, Y. Kanai, T. Kondo, I. Imoto, J. Inazawa & S. Hirohashi, 2005. Genetic classification of lung adenocarcinoma based on array-based comparative genomic hybridization analysis: its association with clinicopathologic features. *Clin Cancer Res,* 11(17), 6177-85.

Sjoblom, T., 2008. Systematic analyses of the cancer genome: lessons learned from sequencing most of the annotated human protein-coding genes. *Curr Opin Oncol,* 20(1), 66-71.

Tonon, G., K. K. Wong, G. Maulik, C. Brennan, B. Feng, Y. Zhang, D. B. Khatry, A. Protopopov, M. J. You, A. J. Aguirre, E. S. Martin, Z. Yang, H. Ji, L. Chin & R. A. Depinho, 2005. High-resolution genomic profiles of human lung cancer. *Proc Natl Acad Sci U S A,* 102(27), 9625-30.

Ullmann, R., S. Petzmann, A. Sharma, P. T. Cagle & H. H. Popper, 2001. Chromosomal aberrations in a series of large-cell neuroendocrine carcinomas: unexpected divergence from small-cell carcinoma of the lung. *Hum Pathol,* 32(10), 1059-63.

Ullmann, R., A. Schwendel, H. Klemen, G. Wolf, I. Petersen & H. H. Popper, 1998. Unbalanced chromosomal aberrations in neuroendocrine lung tumors as detected by comparative genomic hybridization. *Hum Pathol,* 29(10), 1145-9.

Watson, S. K., R. J. deLeeuw, D. E. Horsman, J. A. Squire & W. L. Lam, 2007. Cytogenetically balanced translocations are associated with focal copy number alterations. *Hum Genet,* 120(6), 795-805.

Weir, B. A., M. S. Woo, G. Getz, S. Perner, L. Ding, R. Beroukhim, W. M. Lin, M. A. Province, A. Kraja, L. A. Johnson, K. Shah, M. Sato, R. K. Thomas, J. A. Barletta, I. B. Borecki, S. Broderick, A. C. Chang, D. Y. Chiang, L. R. Chirieac, J. Cho, Y. Fujii, A. F. Gazdar, T. Giordano, H. Greulich, M. Hanna, B. E. Johnson, M. G. Kris, A. Lash, L. Lin, N. Lindeman, E. R. Mardis, J. D. McPherson, J. D. Minna, M. B. Morgan, M. Nadel, M. B. Orringer, J. R. Osborne, B. Ozenberger, A. H. Ramos, J. Robinson, J. A. Roth, V. Rusch, H. Sasaki, F. Shepherd, C. Sougnez, M. R. Spitz, M. S. Tsao, D. Twomey, R. G. Verhaak, G. M. Weinstock, D. A. Wheeler, W. Winckler, A. Yoshizawa, S. Yu, M. F. Zakowski, Q. Zhang, D. G. Beer, Wistuba, II, M. A. Watson, L. A. Garraway, M. Ladanyi, W. D. Travis, W. Pao, M. A. Rubin, S. B. Gabriel, R. A. Gibbs, H. E. Varmus, R. K. Wilson, E. S. Lander & M. Meyerson, 2007. Characterizing the cancer genome in lung adenocarcinoma. *Nature,* 450(7171), 893-8.

Wilson, I. M., J. J. Davies, M. Weber, C. J. Brown, C. E. Alvarez, C. MacAulay, D. Schubeler & W. L. Lam, 2006. Epigenomics: mapping the methylome. *Cell Cycle,* 5(2), 155-8.

6. Defining clonal relationships and evaluating clonal evolution[15]

6.1. Introduction

Resistance to chemotherapy is a major barrier to effective treatment for many cancer types. Resistance mechanisms may either act directly against a drug (e.g. limiting intracellular drug accumulation, increasing drug detoxification, or failing to convert drug precursors into their active form) or act by compensating for drug-induced effects (e.g. altering amounts or activities of drug targets, activating analogous pathways not targeted by drugs, or increasing DNA repair and anti-apoptotic signalling) (Yasui et al. 2004). Where multiple drugs are used to treat a tumor, multidrug resistance (MDR) may arise – sometimes by activation of a single gene (Szakacs et al. 2006). In order to develop effective treatment strategies for the clinic, it is necessary to understand mechanisms of chemoresistance (Leonard et al. 2003).

Cell models such as the NCI-60 cell lines have been the *in vitro* platform for identifying drug resistance factors. Exposure of cell cultures to chemotherapeutic agents has led to the identification of genes that confer resistance by virtue of their increased gene dosage and expression level. Classic examples of gene discovery include P-glycoprotein (*ABCB1*), multidrug resistance-associated protein 1 (*ABCC1*), and dihydrofolate reductase (*DHFR*) (Alt et al. 1978; Cole et al. 1992; Juliano & Ling 1976).

Previously, Bradley et al. characterized DNA copy number and expression changes of *ABCB1* in a series of vincristine-selected SKOV3 ovarian adenocarcinoma cell line derivatives (SKVCRs) (Bradley et al. 1989). These SKVCR lines, derived in series through exposure to increasing amounts of drug, exhibited escalating resistance to a broad variety of therapeutic agents, including vincristine, vinblastine, adriamycin, and colchicine, demonstrating a clear MDR phenotype. Increases in *ABCB1* expression at both mRNA and protein levels were observed with increasing resistance, with

[15] A version of this chapter has been published: Buys TPH, Chari R, Lee EHL, Zhang M, MacAulay C, Lam S, Lam WL, Ling V. (**2007**) "Genetic changes in the evolution of multidrug resistance for cultured human ovarian cancer cells" *Genes, Chromosomes & Cancer*, 46:1069-1079.

concurrent gene amplification detected in the more resistant lines. Significantly, *ABCB1* expression increases were observed prior to gene amplification and increases in ABCB1 protein levels were seen without concurrent spikes in transcript levels, suggesting complex activation of the MDR phenotype. Moreover, low level MDR was observed prior to marked increases in *ABCB1* expression, suggesting alternative mechanisms contributing to the resistance phenotype (Bradley et al. 1989).

In this study we used an integrative approach to track global genomic and gene expression changes that parallel the escalation of the MDR phenotype in SKVCRs. Our results suggest that not only might multiple genes play a role in driving ABCB1-mediated resistance, but also that the MDR phenotype may in fact emerge in these lines through activation of resistance genes that function independently of *ABCB1*.

6.2. Materials and methods

6.2.1. Cell culture, DNA and RNA isolation

Vincristine-resistant derivatives of the ovarian adenocarcinoma cell line SKOV3 (SKVCRs) were previously described (Bradley et al. 1989). No mutagens were used to derive resistant sublines. All cells were routinely maintained in α-MEM medium with 15% fetal calf serum and grown at 37°C in humidified atmosphere containing 5% CO_2. All SKVCR lines were passaged four times in the presence of maximum tolerable drug concentration (at 0, 0.015, 0.10, 0.25 or 2.0 μg/ml vincristine for SKOV3, SKVCR 0.015, SKVCR 0.10, SKVCR 0.25, and SKVCR 2.0, respectively) prior to harvesting for DNA and RNA extraction. Cell pellets were resuspended in a DNA lysis buffer (10 mM Tris-HCl, pH7.5, 50 mM NaCl, 1 mM EDTA, 0.5% SDS) and digested with Proteinase K (Invitrogen) at 55˚C for 48 hours. DNA extraction and RNase-treatment were performed as previously described (Garnis et al., 2003). DNA quantities were determined using an ND-1000 Spectrophotometer (Nanodrop). RNA was extracted from isolated cells using the RNeasy Mini Kit (Qiagen) and re-dissolved in RNase-free water.

6.2.2. Gene expression profiling

The SKOV3 and SKVCR 2.0 cell lines were profiled using the Affymetrix Human

Genome U133 Plus 2.0 microarray at the McGill University Genome Quebec Innovation Centre. Briefly, this array allows the simultaneous analysis of >47,000 different transcripts. The data were then processed and normalized using the Microarray Suite 5.0 (MAS5) algorithm and mean scale normalized prior to further differential analysis, with the "control" group comprised of two samples of SKOV3 and the "treated" group comprised of two samples of SKVCR 2.0. To determine which changes were significant, a standard two fold change between the control group mean and treated group mean was used. In addition, the difference between the control and the treated group means was required to exceed 100, and the call measurement for the given probe had to be present ("P") in at least 1 of the 4 samples. 5,569 probesets met these stringent criteria (see Supplementary Table 6.1)[16]. Heat maps of log2 transformed gene expression array data were generated using Genesis software (Sturn et al. 2002). Gene expression microarray results have been made publicly available at the NCBI Gene Expression Omnibus (Accession No. GSE7556).

6.2.3. Whole genome tiling-path array CGH

Segmental DNA copy number status was assessed for the SKOV3 genome and the genomes of its SKVCR derivative lines by array CGH. Tiling-path CGH arrays were manufactured at the British Columbia Cancer Research Centre. Each array is comprised of >32,000 overlapping BAC clones, spotted in triplicate, that span the entire human genome (Ishkanian et al. 2004). Array CGH experiments were undertaken as previously described (Garnis et al. 2004). 400 ng each of test and reference genomic DNA (Novagen) were differentially labelled with cyanine-5 and cyanine-3 dye respectively, using a random priming protocol. Unincorporated nucleotides were removed using ProbeQuant Sephadex G-50 Columns (Amersham). The sample mixture was co-precipitated with 200 µg of C_0t-1 DNA (Invitrogen), resuspended in a 90 µl cocktail consisting of DIG Easy hybridization solution (Roche), sheared herring sperm DNA (Sigma-Aldrich), and yeast tRNA (Calbiochem), denatured at 85˚C for 10 min, followed by a 1 hour incubation at 45˚C, and hybridized to the array at 45˚C for 36

[16] All supplementary data from this manuscript may be found at the journal website (http://www3.interscience.wiley.com/journal/38250/home).

hours. Slides were washed in 0.1X saline sodium citrate (SSC), 0.1% SDS solution at room temperature five times for 5 min each, rinsed with 0.1X SSC, and then dried by centrifugation. Arrays were imaged using a CCD-based imaging system (Arrayworx eAuto, Applied Precision) and analyzed using SoftWoRx Tracker Spot Analysis software. A normalization framework was applied to detect and correct any systematic platform bias (Khojasteh et al. 2005). Spot data were visualized in a karyogram format using SeeGH software (Chi et al. 2004). Data filtering and breakpoint identification was performed as previously described, with replicate clones having standard deviations >0.075 and signal-to-noise ratios <3 excluded from the raw data (Baldwin et al. 2005). Alteration status for each BAC clone and genomic breakpoint boundaries were defined by aCGH-Smooth software (Lambda = 6.75, Maximum number of breakpoints per chromosome = 100, and Minimum difference between levels = 0.075. Default program settings were used for all other parameters.) (Coe et al. 2006; Jong et al. 2004). Clones with non-informative values (i.e. those filtered out with the above criteria) had their alteration status inferred by neighbouring clones. Array CGH data for all profiled samples have been made publicly available through the System for Integrative Genomic Microarray Analysis (SIGMA) on-line resource (http://sigma.bccrc.ca) (Chari & Lockwood et al. 2006).

6.2.4. Semi-quantitative reverse-transcriptase-PCR

Expression levels of genes of interest were determined by reverse-transcriptase PCR (RT-PCR) using gene-specific primers. cDNA was synthesized for SKOV3 and all SKVCR sublines using SuperScript II Reverse Transcriptase (Invitrogen). Table 6.1 summarizes the sequences of primers used in this study. PCR cycle conditions were as follows: denaturation for 1 min at 95°C, followed by 30−35 cycles of (30 seconds at 95°C, 30 seconds at specified annealing temperature for a given gene, 1 min at 72°C), and a 10 min extension at 72°C. RT-PCR products were resolved by polyacrylamide gel electrophoresis, imaged by SYBR green staining (Roche) on a Molecular Dynamics Storm Phosphoimager model 860 and quantified using ImageQuant software (Molecular Dynamics). β-actin levels were used for normalization between samples.

6.2.5. Analysis of pathway activation

Gene lists were generated based on differential analysis of genomic and gene expression data from SKOV3 and SKVCR cell lines. Functional annotation and differential pathway activation between drug sensitive (SKOV3) and highly resistant (SKVCR 2.0) lines was determined for genes identified in Supplementary Table 6.1 with Ingenuity Pathway Analysis software (Ingenuity). Genomic status in SKVCR 2.0, defined by the segmented array CGH data, was then established for activated pathway members in order to determine the contribution of alterations at the DNA level to the observed gene expression changes.

6.3. Results

6.3.1. High level genomic and gene expression differences between parental line and MDR line SKVCR 2.0

Comparison of the parental vincristine-sensitive SKOV3 and the drug resistant SKVCR 2.0 gene expression profiles using the Affymetrix U133 2.0 Plus platform revealed striking gene expression differences (Supplementary Table 6.1). The top 25 genes overexpressed with resistance are listed in Figure 6.1. *ABCB1* exhibited a marked increase in transcript level for SKVCR 2.0, which is consistent with the pivotal role of this gene in the MDR phenotype. Figure 6.1 shows that one third of genes highly over-expressed in SKVCR 2.0 were ECM-associated factors, including *FN1* (the most differentially expressed gene observed), *COL1A2*, *COL5A1*, *COL8A1, FLRT3*, *HAS2*, *CSPG2*, and *THBS1*.

To investigate if changes in gene dosage were a mechanism for the observed overexpression, we compared the genomes of SKOV3 and SKVCR 2.0 by whole genome tiling-path array CGH. Direct competitive hybridization of DNA samples of these two lines revealed overall similarity in genomic profile, but also showed distinct segmental DNA copy number differences (Figure 6.2). Both broad alterations spanning multiple chromosome bands and focal sub-megabase alterations were detected. *ABCB1* gene amplification was evident, with the identification of a high level amplicon of

<700 kb at 7q21.12. A second ABC transporter gene, *ABCB4*, was amongst the seven genes within this amplicon. Examination of gene expression microarray data showed that *ABCB4* had elevated expression with increased resistance, though this difference between SKOV3 and SKVCR 2.0 was less than the observed difference in *ABCB1* expression (Supplementary Table 6.1). Besides *ABCB1*, of the differentially expressed genes identified in Figure 6.1, *CSPG2*, *THBS1*, *CDH13*, and *CYBA* showed DNA copy number gains with concurrent overexpression (Figure 6.2).

6.3.2. Global genomic analysis to identify disruption in gene networks

We next performed an unbiased global analysis of expression data for all genes defined as differentially expressed between SKOV3 and SKVCR 2.0 (Supplementary Table 6.1). Functional annotation of these genes and evaluation of gene network disruption – using IPA software to assess concerted changes – revealed that the most highly activated network in SKVCR 2.0 was comprised of ECM-associated factors. Specifically, identified genes included those involved with ECM-mediated activation of tumor cell invasion and cell proliferation processes (Table 6.2, Figure 6.3) (Adamia et al. 2005; Poon et al. 2004; Tringler et al. 2005; Wu et al. 2005).

We then examined if there were genetic bases for the disruption of this gene network. Segmental genomic gains and losses were defined for the SKVCR 2.0 versus SKOV3 tiling-path array CGH profile by the *aCGH-Smooth* algorithm (Coe et al. 2006; Jong et al. 2004). All genomic data were then integrated with expression data for those genes defined as differentially expressed between SKVCR 2.0 and SKOV3 (Supplementary Table 6.2).

Our data show a correlation for genomic alterations in activating the ECM factors described above, with approximately a third of the overexpressed ECM genes showing increased DNA copy number in SKVCR 2.0 (Figure 6.2 and 6.3).

6.3.3. Characterization of genomic alterations in the escalation of resistance by analyzing intermediate lines

DNA copy number status throughout the genomes of SKOV3 and its derivative lines of

increasing resistance (SKVCR 0.015, SKVCR 0.10, SKVCR 0.25, and SKVCR 2.0 – see Materials and Methods) were assessed by tiling-path array CGH. Approximately 5.6% (5.5% gains, 0.1% losses) of genomic alterations represented by BAC clones were shared in all SKVCR lines and absent in SKOV3 (Supplementary Table 6.3). On average, over two hundred segmental DNA copy number alterations were observed for each cell line (Supplementary Table 6.4).

Seven of the ECM-associated genes described above were shown to reside within segmental DNA gains across all SKVCR lines (Figure 6.3). Gene expression analysis of five representative components of the ECM gene network by semi-quantitative RT-PCR across all cell lines revealed overexpression at even the lowest level of resistance (Figure 6.4). The expression levels of *FN1*, *HAS2*, *THBS1*, *CSPG2*, and *CD44* were increased in SKVCR 0.015 compared to the SKOV3 parental line. Greater expression of these genes was observed across all the SKVCR derivative lines.

Significantly, we detected novel genomic alterations present only in SKVCR lines with intermediate resistance levels. The most conspicuous of these alterations was a 1.79 Mb amplicon at 16p13 present in SKVCR 0.015 and SKVCR 0.10 that encompassed the ABC transporter genes *ABCC1* and *ABCC6* (Figure 6.5A). The level of this segmental copy number alteration was the highest detected for any lines in this study. Semi-quantitative RT-PCR experiments confirmed the functional consequence of this genetic event as these transporter genes were correspondingly overexpressed in SKVCR 0.015 and SKVCR 0.10 cell populations relative to parental SKOV3, but not in the more resistant SKVCR 0.25 and SKVCR 2.0 lines, which no longer contained evidence of this amplicon (Figure 6.5C). In contrast, *ABCB1* (and *ABCB4*) expression escalated with increasing resistance, with high levels of expression observed for SKVCR 0.25 and SKVCR 2.0, both of which harboured a 0.66 Mb amplicon at 7q21.12 (Figures 6.5B and 6.5C). Only minimal *ABCB1* expression was observed for SKOV3 and for the lines exhibiting “intermediate” levels of MDR (SKVCR 0.015 and SKVCR 0.10) (Supplementary Figure 6.2). This was consistent with previous reports of *ABCB1* expression for these lines (Bradley et al. 1989).

6.4. Discussion

6.4.1. Identification of factors complementing ABCB1-mediated MDR

Because ABCB1 activation was previously described as the major driving force behind resistance in the SKVCR lines (Bradley et al. 1989), it is anticipated that *ABCB1* would be one of the most differentially expressed gene between these lines (Figure 6.1). Notably, recent studies have implicated the contribution of ECM-associated genes in the MDR phenotype (Adamia et al. 2005; Miletti-Gonzalez et al. 2005; Misra et al. 2005; Poon et al. 2004; Shain & Dalton 2001; Toole 2004; Tringler et al. 2005; Wu et al. 2005). For example, it has been shown that CD44 and hyaluronan (product of the *HAS2* gene) activate ABCB1, with overexpression of these genes demonstrated to directly drive resistance mechanisms *in vitro* (Miletti-Gonzalez et al. 2005; Misra et al. 2005; Toole 2004). Overexpression of many ECM-associated genes is seen in our results. It is possible that these ABCB1-activating factors in SKVCR 2.0 may act synergistically with *ABCB1* amplification to drive MDR in response to the extremely high drug concentrations faced by SKVCR 2.0. Furthermore, additional factors known to play an upstream role in invasion and/or activation of CD44/hyaluronan signalling were detected by our gene expression profiling analysis to be overexpressed in SKVCR 2.0 (Figure 6.3) (Adamia et al. 2005; Poon et al. 2004; Ricciardelli & Rodgers 2006; Tringler et al. 2005; Wu et al. 2005). Given the pronounced levels of expression increase we observe for these factors, and their reported interactions with CD44 and hyaluronan, these genes may also play a role in the activation of ABCB1-mediated MDR. Interestingly, while high level gene amplification and overexpression appears to be the process for activation of *ABCB1*, our integrated genomic and gene expression data suggest that activation of the ECM-associated factors described above is characterized by more subtle increases in gene dosage and expression.

6.4.2. Activation of ECM and invasion genes with emerging resistance

FN1 exhibits a significant change in expression between the vincristine-sensitive SKOV3 and the resistant derivative SKVCR 2.0 (Figure 6.1). It has previously been demonstrated that FN1 may contribute to drug resistance through cell adhesion-

mediated drug resistance (CAM-DR) and that this phenotype is essential for cell survival at the time of initial drug exposure (Shain & Dalton 2001; Hazlehurst et al. 2006). The increase in *FN1* expression we detected at even the lowest levels of resistance (SKVCR 0.015) (Figure 6.4) may suggest a role for *FN1* in cell survival prior to the establishment of resistance through *ABCB1* amplification. In addition, we also observed overexpression of other ECM factors such as *CSPG2* beginning at the lowest level of resistance (SKVCR 0.015) and continuing through the subsequent stages (Figure 6.4). Moreover, our data showed genomic activation in this ECM-associated pathway at all levels of resistance, including increased DNA copy number for *FBLN2*, *ITGA3*, *MMP2*, *MMP15*, *MMP24*, *TIMP2*, and *TNC* across all SKVCR lines (Figure 6.3). Although a direct role of ECM disruption in drug resistance in for the SKVCR lines is unclear, these genes have previously been reported in CAM-DR-based survival in response to drug selection (Cattaruzza et al. 2004; Zheng et al. 2004).

6.4.3. Succession in ABC transporter-mediated resistance

Multiple ABC transporter genes have previously been implicated in drug resistance for tumor cells. Given that amplification of the ABC transporter *ABCB1* had been described as contributing to MDR in the SKVCR cell lines, we decided to simultaneously track all known ABC transporter genes for DNA copy number increase across all SKVCRs (Borst & Elferink 2002; Frank et al. 2003; Stefkova et al. 2004; Uitto 2005). Besides *ABCB1*, *ABCB4*, *ABCC1*, and *ABCC6*, no high level copy number changes were observed (Supplementary Figure 6.1). As stated above, *ABCB4* was a part of the amplicon at 7q21.12 encompassing *ABCB1* and exhibited increases in gene expression that paralleled the behavior of its neighbor (Figure 6.5B and 6.5C). However, *ABCB4* has previously been shown to have minimal importance in resistance to mitotic inhibitors (Duan et al. 2004), suggesting that it may in fact be a passenger alteration that arises through selection for ABCB1 activation.

We also observed a prominent high level amplification on chromosome 16 found in both SKVCR 0.015 and SKVCR 0.10 that encompassed multiple genes, including two ABC transporters: *ABCC1* and *ABCC6* (Figure 6.5A). This was the largest genomic

alteration observed for any of the cell lines in this study. *ABCC1* has been implicated in poor prognosis and drug resistance for numerous cancer types and numerous chemotherapeutic agents (Kruh & Belinsky 2003). While *ABCC6* has been shown to transport organic anions and play a role in the heritable disorder pseudoxanthoma elasticum (Uitto, 2005), its role in drug resistance is still being determined. Although earlier work suggests that *ABCC6* activation is a by-product of drug selection for *ABCC1* amplification (Kool et al. 1999), the discovery of drug substrates for ABCC6 bolsters its role as a potential mediator of the MDR phenotype (Belinsky et al. 2002; Kruh & Belinsky 2003). Interestingly, *ABCC6* has previously been implicated in "intermediate" level drug resistance (Schmidt et al. 2003; Schmidt et al. 2004), which is consistent with our findings in the SKVCR 0.015 and SKVCR 0.10 lines. Semi-quantitative RT-PCR clearly showed overexpression of *ABCC1* and *ABCC6* (Figure 6.5C), confirming the biological significance of the observed amplifications of these transporter genes.

The fact that the boundaries of this amplicon were identical between SKVCR 0.015 and SKVCR 0.10 suggests that it did not arise independently in these lines, indicating that this genetic alteration was preserved at and likely important to "intermediate" level MDR. (The increased magnitude of this amplicon in SKVCR 0.10 may have been the result of further amplification of the alteration already present in SKVCR 0.015.) The fact that this alteration was not detected in the lines exhibiting higher levels of drug resistance (SKVCR 0.25 and SKVCR 2.0), indicates that an alternative process – likely *ABCB1* activation – was driving MDR at higher drug dosage. This suggests that a succession of amplification-mediated mechanisms may regulate the MDR phenotype (specifically, amplifications spanning different ABC transporters).

In conclusion, our data show that multiple genes – including different ABC transporter genes – are activated with escalating vincristine resistance in SKOV3, suggesting that the activation of the MDR phenotype is complex and multi-faceted. This is consistent with earlier findings, including previous analysis of paclitaxel resistance for ovarian cancer (Duan et al. 2006). Our finding that a specific segmental DNA alteration

containing *ABCC1* and *ABCC6* is amplified and later displaced by an amplicon spanning *ABCB1* suggests that a population of cells is capable of evolving successive drug resistance mechanisms. In light of this, therapeutic measures to address MDR as a single-gene trait may need to be re-evaluated. Moreover, the sequential appearance of these amplicons suggests that it may be useful to consider longitudinal monitoring of genetic changes as part of disease management strategies for ovarian and other cancer types. Without knowledge of alternate avenues to MDR, many cancer types will continue to be intractable.

Table 6.1 – Semi-quantitative RT-PCR conditions and gene primers.

Gene name	Sequence serving as basis for primers	Forward primer	Reverse primer	$MgCl_2$ (mM)	Cycles	Annealing T_m (°C)
ABCB1	NM_000927	ttgtccaaactgcctgtgaa	ccaagaagaatgaagccaga	2	32	58
ABCB4	NM_018850	cctcaaatcctcctgttgga	gtgccatgctccttgactct	1	45	56
HAS2	NM_005328	cctggatctcattcctcagc	gagatgcctgtcatcaccaa	1	35	56
THBS1	NM_003246	gcaaagaaagccatgaggtc	ctctttccgcttgatccaaa	1	45	52
FN1	NM_002026	ggtgccatgacaatggtgt	gcaaatggcaccgagatatt	1	45	58
CSPG2	NM_004385	ccagtgcagtccatttcctaa	agcagccgaaccaatgatta	1	35	58
CD44	NM_000610	cattggccagatccattttc	ggagtggcttgttgcttttc	1	45	51
ß-actin	NM_001101	gatgtggatcagcaagca	gaaagggtgtaacgcaact	2	35	55
ABCC1	NM_004996	gaagggacctggttggacta	catcaatcatggtgggatca	2	45	58
ABCC6	NM_001171	agatccacgcaggagagaag	ggatgatgctgatcctggag	1	40	58

Figure 6.1

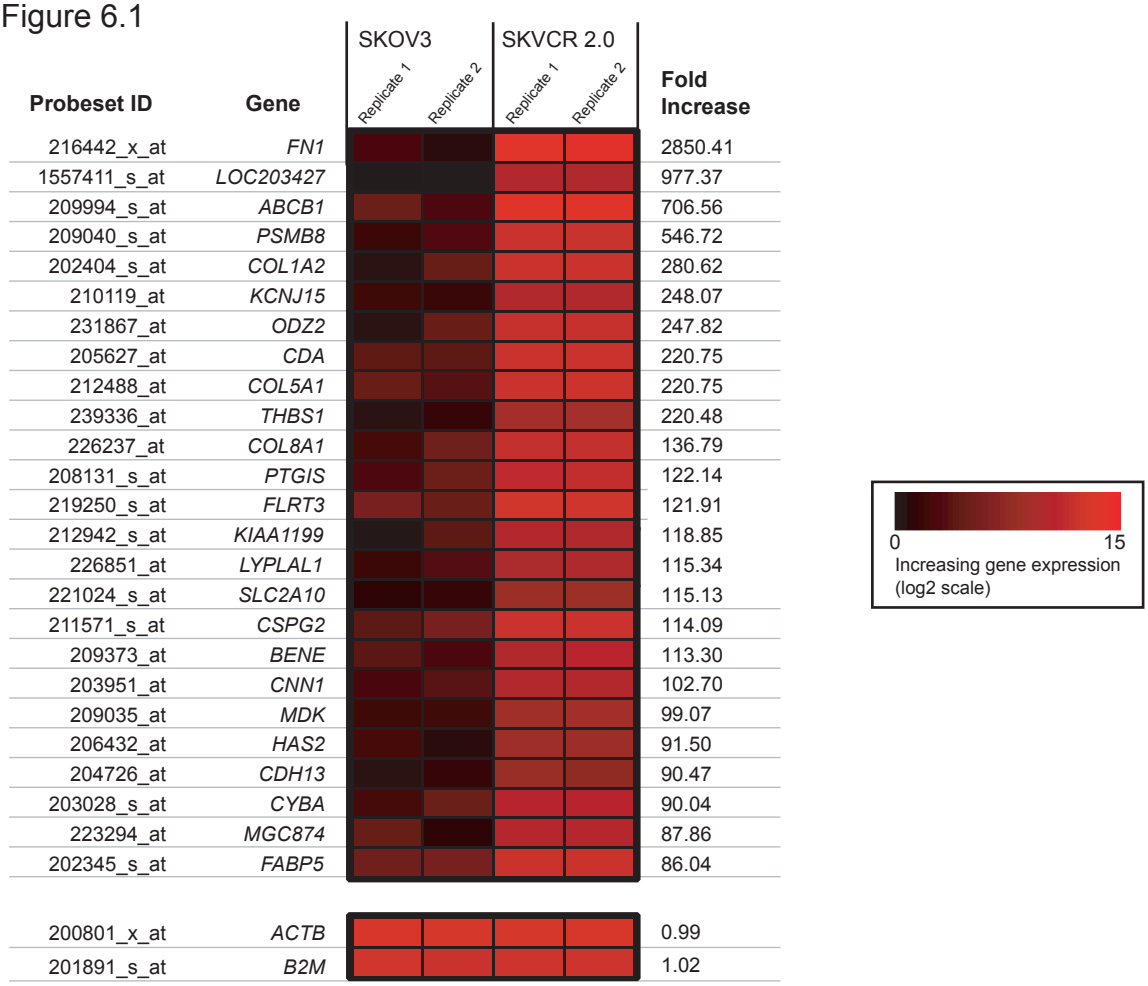


Figure 6.1 – Top 25 genes increased in expression for SKVCR 2.0 relative to SKOV3. Data were from Affymetrix expression microarray profiling. Expression level and fold change values were calculated for genes based on representative probesets listed in this table. Genesis software was used to create colorimetric representations of gene expression data, with greater intensity of the red color denoting increased expression. Gene expression data for the housekeeping genes β-actin (ACTB) and β-2-microglobulin (B2M), which showed equivalent expression in all experiments, were also included. All expression data included in the heat map were log2 transformed.

Figure 6.5

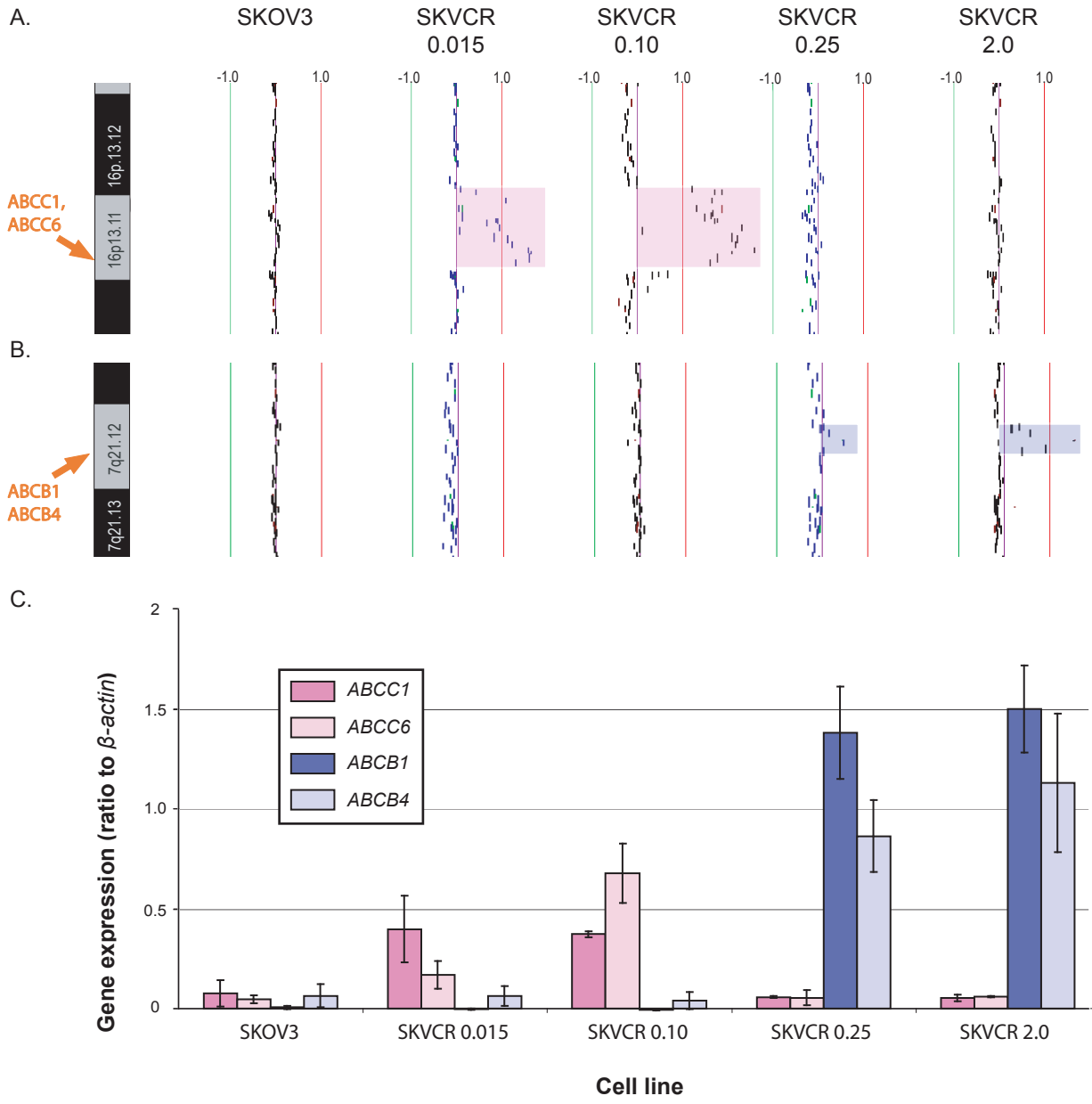

A.
SKOV3
SKVCR 0.015
SKVCR 0.10
SKVCR 0.25
SKVCR 2.0
-1.0
1.0
ABCC1,
ABCC6
16p13.12
16p13.11
B.
ABCB1
ABCB4
7q21.12
7q21.13
C.
Gene expression (ratio to β-actin)
2
1.5
1.0
0.5
0
ABCC1
ABCC6
ABCB1
ABCB4
SKOV3
SKVCR 0.015
SKVCR 0.10
SKVCR 0.25
SKVCR 2.0
Cell line

Figure 6.2 – Whole genome SeeGH karyogram of array comparative genomic hybridization of SKOV3 vs. SKVCR 2.0. Each BAC clone on the tiling-path array CGH platform is represented as a black spot at its known chromosomal position. Clones are plotted along the horizontal axis based on the log2 value of their signal intensity ratios from the array experiment. Vertical lines show log2 signal intensity ratios from -1 (green line) to 1 (red line), with copy number increases to the right and decreases to the left of zero (purple line). A log2 signal ratio of zero represents equivalent copy number between the hybridized samples. Where log2 > 0, it suggests increased DNA copy number for SKVCR 2.0 relative to parental SKOV3. Where log2 < 0, the reverse is inferred. Vertical orange bars span alterations at *ABCB1*, *THBS1*, *CSPG2*, *CDH13*, and *CYBA* with higher magnification images of these specific regions also provided (regions of genomic gain within zoomed in image are highlighted in yellow).

Table 6.2 – Genes assessed for pathway activation in SKVCR 2.0.

Gene Symbol	Gene Name	Potential Oncogenic Functions
ADAMTS1	A disintegrin-like and metalloproteinase with thrombospondin type 1 motif, 1	Proliferation
CD44	CD44 antigen	Migration, adhesion, proliferation, apoptosis, invasion, growth
COL18A1	Collagen, type XVIII, alpha-1	Migration, proliferation, apoptosis, adhesion, invasion, growth
CSPG2	Versican	Adhesion, apoptosis, attachement, growth, proliferation
EGR1	Early growth response 1	Apoptosis, growth, differentiation, proliferation, adhesion, transformation
ERCC1	Excision-repair, complementing defective, in chinese hamster, 1	Polyploidization, DNA repair
F3	Coagulation Factor III	Migration, adhesion, apoptosis, cell death
FBLN1	Fibulin 1	Invasion
FBLN2	Fibulin 2	Invasion
FBN1	Fibrillin 1	Cell spreading, growth
FN1	Fibronectin 1	Migration, adhesion, cell spreading, apoptosis, proliferation, survival
FOS	FBJ osteosarcoma virus	Transformation, apoptosis, growth, proliferation, cell death, motility, differentiation, cell cycle progression
FOSL2	FOS-related antigen 2	Differentiation, cell death, proliferation, cell cycle progression
FSTL1	Follistatin-like 1	Invasion
HAS2	Hyaluronan synthase 2	Invasion, growth, migration, transformation
ITGA3	Integrin, alpha 3	Adhesion, migration, attachment, cell spreading, survival, motility, proliferation, differentiation
ITGB3	Integrin, beta 3	Adhesion, migration, attachment, proliferation, cell spreading, differentiation, invasion, survival
KCNIP3	Potassium channel-interacting protein 3	Proliferation
LAMA4	Laminin, alpha 4	Migration, branching, adhesion, binding, alignment, development, elongation, proliferation, degeneration
LAMP2	Lysosome-associated membrane protein 2	Attachment
LOX	Lysyl oxidase	Growth, transformation, proliferation, attachment, invasion
LTBP1	Latent transforming growth factor-beta-binding protein 1	Proliferation, differentiation, growth
MMP15	Matrix metalloproteinase 15	Invasion, growth, proliferation, adhesion, invasiveness, survival, differentiation
MMP2	Matrix metalloproteinase 2	Invasion, migration, growth, proliferation, invasiveness, differentiation
MMP24	Matrix metalloproteinase 24	Invasion, growth, proliferation, malignancy, cell movement, invasiveness, survival, differentiation
NID2	Nidogen 2	Adhesion
PLAUR	Plasminogen activator receptor, urokinase-type	Migration, adhesion, invasion, proliferation, growth, survival
SDC1	Syndecan 1	Cell spreading, migration, adhesion, proliferation, invasion
SERPINE2	Protease inhibitor 7	Growth, migration, proliferation, apoptosis
TGFB1	Transforming growth factor beta 1	Apoptosis, growth, proliferation, differentiation, cell death, cell cycle progression, adhesion, migration
TGFBI	Transforming growth factor beta-induced	Adhesion, apoptosis, proliferation, cell spreading, migration, attachment
TGFBR2	Transforming growth factor beta receptor, type II	Proliferation, growth, differentiation, apoptosis, ploidy
THBS1	Thrombospondin 1	Proliferation, migration, adhesion, cell spreading, activation, cell death
TIMP2	Tissue inhibitor of metalloproteinase 2	Migration, proliferation, growth, invasion, apoptosis, angiogenesis, attachment
TNC	Tenascin C	Adhesion, apoptosis, proliferation, cell spreading, migration, differentiation

Figure 6.3 – Pathway activation of across SKVCR lines. All depicted genes were overexpressed in SKVCR 2.0 relative to drug sensitive SKOV3. The degree of overexpression is represented by intensity of red within the shape representing a given gene. Those genes surrounded by a blue border also have increased DNA copy number in SKVCR 2.0 relative to SKOV3. Those genes with black-dashed border exhibited increased copy number across all SKVCR lines. Where green lines exist between genes, an activating relationship is indicated (with arrow depicting direction of activation). Black lines between genes show known interactions. Gene relationships and figure layout are based on Ingenuity Pathway Analysis and references provided elsewhere in text.

Figure 6.3

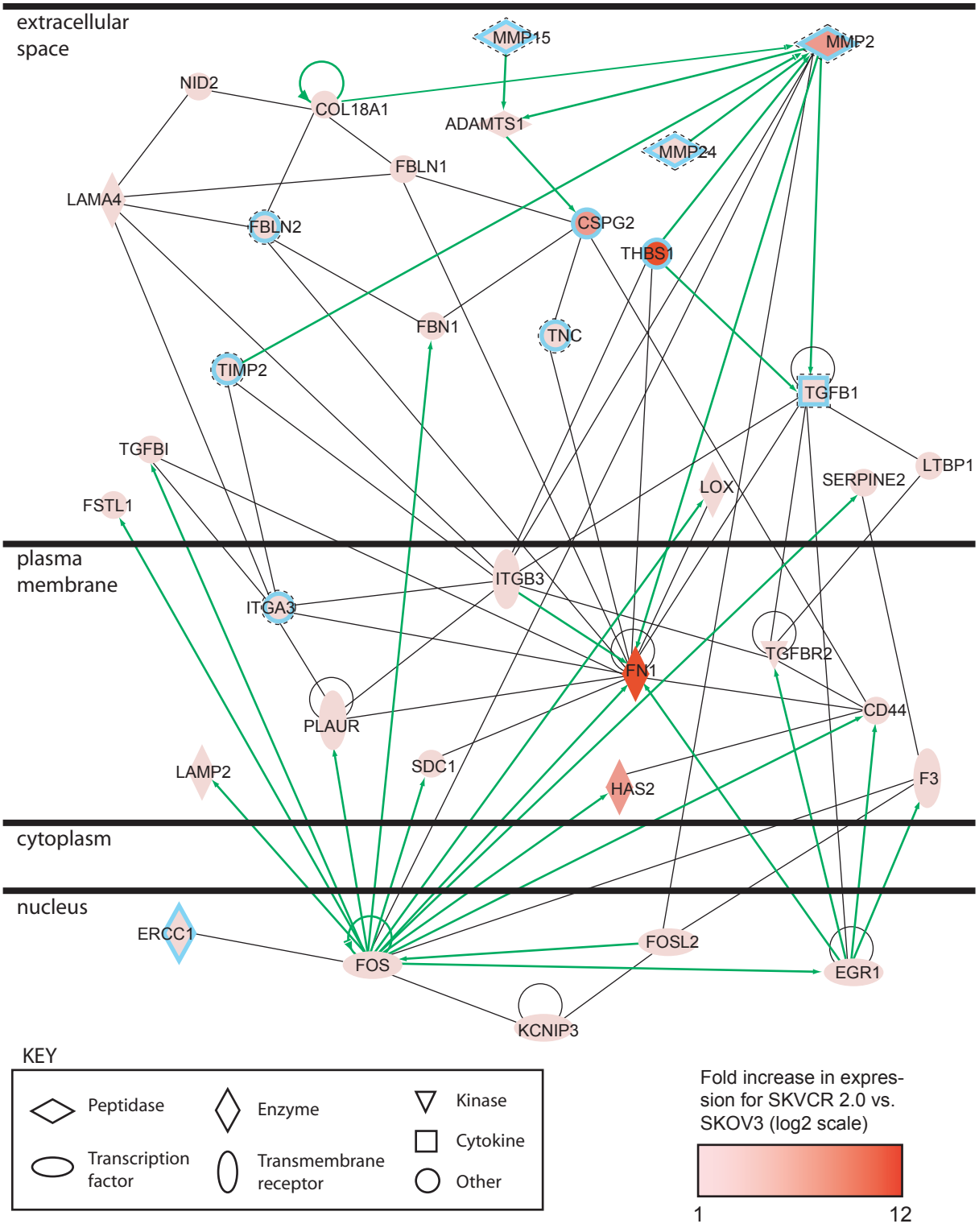

extracellular space
plasma membrane
cytoplasm
nucleus
MMP15
MMP2
NID2
COL18A1
ADAMTS1
MMP24
FBLN1
LAMA4
FBLN2
CSPG2
THBS1
FBN1
TNC
TIMP2
TGFB1
TGFBI
LTBP1
SERPINE2
LOX
FSTL1
ITGB3
ITGA3
TGFBR2
FN1
CD44
PLAUR
LAMP2
SDC1
HAS2
F3
ERCC1
FOSL2
FOS
EGR1
KCNIP3
KEY
Peptidase
Enzyme
Kinase
Transcription factor
Transmembrane receptor
Cytokine
Other
Fold increase in expression for SKVCR 2.0 vs. SKOV3 (log2 scale)
1
12

Figure 6.4

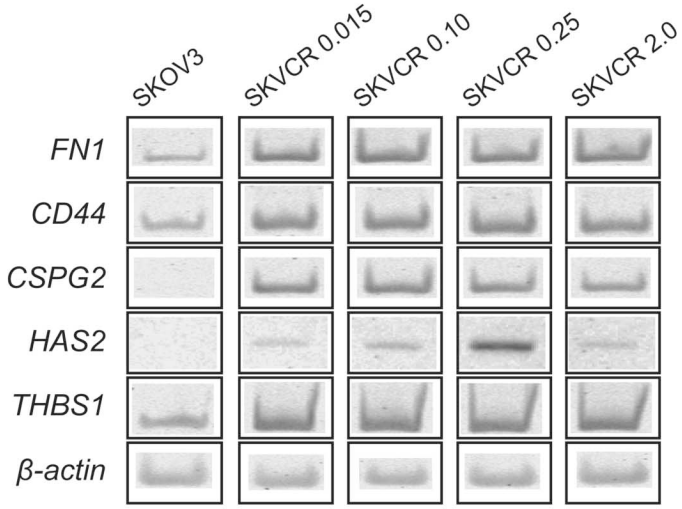
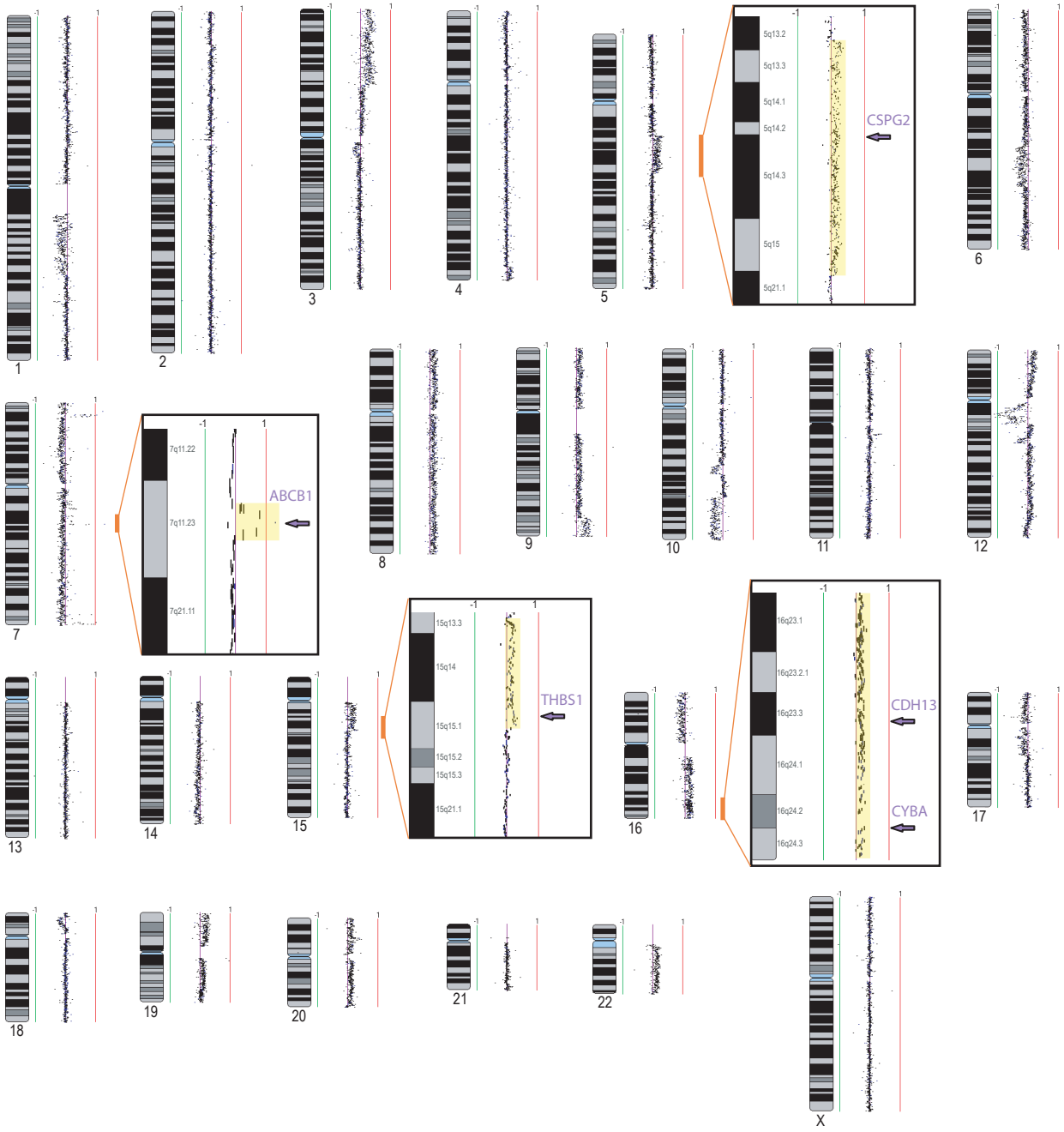


Figure 6.4 – Expression levels of ECM and tumor invasion genes. Semi-quantitative RT-PCR analysis of HAS2, CD44, CSPG2, FN1, THBS1 and β-actin expression in SKOV3 and the SKVCR derivatives. Representative results from replicate experiments are shown.

Figure 6.5 – Assessment of genetic alteration and gene expression changes at 16p13 and 7q12. A. Tiling-path array CGH profiles of 16p13.13-p12.3 containing *ABCC1*and *ABCC6* genes. The SeeGH karyogram was annotated as described in Figure 6.2. Each cell line sample was hybridized against the same reference DNA sample. The amplicons in the SKVCR 0.015 and SKVCR 0.10 lines are labelled in pink. B. Tiling-path array CGH profiles of 7q12.11-q12.13 containing *ABCB1* and *ABCB4*. Regions of amplification are highlighted in blue. C. Results for replicate semi-quantitative RT-PCR experiments for *ABCC1, ABCC6, ABCB1*, and *ABCB4*. Expression levels for each gene were given as ratios to β-actin expression, with SE for each group also shown. *ABCB1* expression levels in SKOV3, SKVCR 0.015, and SKVCR 0.10 were too low for detection under the conditions of the RT-PCR experiment. Additional real-time quantitative PCR experiments confirmed low level *ABCB1* expression for these lines, consistent with earlier findings (Supplementary Figure 6.2, [Bradley et al. 1989]).

Figure 6.2


1
2
3
4
5
5q13.2
5q13.3
5q14.1
5q14.2
5q14.3
5q15
5q21.1
CSPG2
6
7
7q11.22
7q11.23
7q21.11
ABCB1
8
9
10
11
12
13
14
15
15q13.3
15q14
15q15.1
15q15.2
15q15.3
15q21.1
THBS1
16
16q23.1
16q23.2.1
16q23.3
16q24.1
16q24.2
16q24.3
CDH13
CYBA
17
18
19
20
21
22
X

6.5. References

Adamia, S., C. A. Maxwell & L. M. Pilarski, 2005. Hyaluronan and hyaluronan synthases: potential therapeutic targets in cancer. *Curr Drug Targets Cardiovasc Haematol Disord,* 5(1), 3-14.

Alt, F. W., R. E. Kellems, J. R. Bertino & R. T. Schimke, 1978. Selective multiplication of dihydrofolate reductase genes in methotrexate-resistant variants of cultured murine cells. *J Biol Chem,* 253(5), 1357-70.

Baldwin, C., C. Garnis, L. Zhang, M. P. Rosin & W. L. Lam, 2005. Multiple microalterations detected at high frequency in oral cancer. *Cancer Res,* 65(17), 7561-7.

Belinsky, M. G., Z. S. Chen, I. Shchaveleva, H. Zeng & G. D. Kruh, 2002. Characterization of the drug resistance and transport properties of multidrug resistance protein 6 (MRP6, ABCC6). *Cancer Res,* 62(21), 6172-7.

Borst, P. & R. O. Elferink, 2002. Mammalian ABC transporters in health and disease. *Annu Rev Biochem,* 71, 537-92.

Bradley, G., M. Naik & V. Ling, 1989. P-glycoprotein expression in multidrug-resistant human ovarian carcinoma cell lines. *Cancer Res,* 49(10), 2790-6.

Cattaruzza, S., M. Schiappacassi, K. Kimata, A. Colombatti & R. Perris, 2004. The globular domains of PG-M/versican modulate the proliferation-apoptosis equilibrium and invasive capabilities of tumor cells. *Faseb J,* 18(6), 779-81.

Chari, R., W. W. Lockwood & W. L. Lam, 2006. Computional Methods for the Analysis of Array Comparative Genomic Hybridization. *Cancer Informatics,* 2, 48-58.

Chi, B., R. J. DeLeeuw, B. P. Coe, C. MacAulay & W. L. Lam, 2004. SeeGH--a software tool for visualization of whole genome array comparative genomic hybridization data. *BMC Bioinformatics,* 5(1), 13.

Coe, B. P., E. H. Lee, B. Chi, L. Girard, J. D. Minna, A. F. Gazdar, S. Lam, C. MacAulay & W. L. Lam, 2006. Gain of a region on 7p22.3, containing MAD1L1, is the most frequent event in small-cell lung cancer cell lines. *Genes Chromosomes Cancer,* 45(1), 11-9.

Cole, S. P., G. Bhardwaj, J. H. Gerlach, J. E. Mackie, C. E. Grant, K. C. Almquist, A. J. Stewart, E. U. Kurz, A. M. Duncan & R. G. Deeley, 1992. Overexpression of a

transporter gene in a multidrug-resistant human lung cancer cell line. *Science,* 258(5088), 1650-4.

Duan, Z., K. A. Brakora & M. V. Seiden, 2004. Inhibition of ABCB1 (MDR1) and ABCB4 (MDR3) expression by small interfering RNA and reversal of paclitaxel resistance in human ovarian cancer cells. *Mol Cancer Ther,* 3(7), 833-8.

Duan, Z., R. Foster, K. A. Brakora, R. Z. Yusuf & M. V. Seiden, 2006. GBP1 overexpression is associated with a paclitaxel resistance phenotype. *Cancer Chemother Pharmacol,* 57(1), 25-33.

Frank, N. Y., S. S. Pendse, P. H. Lapchak, A. Margaryan, D. Shlain, C. Doeing, M. H. Sayegh & M. H. Frank, 2003. Regulation of progenitor cell fusion by ABCB5 P-glycoprotein, a novel human ATP-binding cassette transporter. *J Biol Chem,* 278(47), 47156-65.

Garnis, C., B. P. Coe, A. Ishkanian, L. Zhang, M. P. Rosin & W. L. Lam, 2004. Novel regions of amplification on 8q distinct from the MYC locus and frequently altered in oral dysplasia and cancer. *Genes Chromosomes Cancer,* 39(1), 93-8.

Hazlehurst, L. A., R. F. Argilagos, M. Emmons, D. Boulware, C. A. Beam, D. M. Sullivan & W. S. Dalton, 2006. Cell adhesion to fibronectin (CAM-DR) influences acquired mitoxantrone resistance in U937 cells. *Cancer Res,* 66(4), 2338-45.

Ishkanian, A. S., C. A. Malloff, S. K. Watson, R. J. DeLeeuw, B. Chi, B. P. Coe, A. Snijders, D. G. Albertson, D. Pinkel, M. A. Marra, V. Ling, C. MacAulay & W. L. Lam, 2004. A tiling resolution DNA microarray with complete coverage of the human genome. *Nat Genet,* 36(3), 299-303.

Jong, K., E. Marchiori, G. Meijer, A. V. Vaart & B. Ylstra, 2004. Breakpoint identification and smoothing of array comparative genomic hybridization data. *Bioinformatics,* 20(18), 3636-7.

Juliano, R. L. & V. Ling, 1976. A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants. *Biochim Biophys Acta,* 455(1), 152-62.

Khojasteh, M., W. L. Lam, R. K. Ward & C. MacAulay, 2005. A stepwise framework for the normalization of array CGH data. *BMC Bioinformatics,* 6, 274.

Kool, M., M. van der Linden, M. de Haas, F. Baas & P. Borst, 1999. Expression of human MRP6, a homologue of the multidrug resistance protein gene MRP1, in tissues and cancer cells. *Cancer Res,* 59(1), 175-82.

Kruh, G. D. & M. G. Belinsky, 2003. The MRP family of drug efflux pumps. *Oncogene,* 22(47), 7537-52.

Leonard, G. D., T. Fojo & S. E. Bates, 2003. The role of ABC transporters in clinical practice. *Oncologist,* 8(5), 411-24.

Miletti-Gonzalez, K. E., S. Chen, N. Muthukumaran, G. N. Saglimbeni, X. Wu, J. Yang, K. Apolito, W. J. Shih, W. N. Hait & L. Rodriguez-Rodriguez, 2005. The CD44 receptor interacts with P-glycoprotein to promote cell migration and invasion in cancer. *Cancer Res,* 65(15), 6660-7.

Misra, S., S. Ghatak & B. P. Toole, 2005. Regulation of MDR1 expression and drug resistance by a positive feedback loop involving hyaluronan, phosphoinositide 3-kinase, and ErbB2. *J Biol Chem,* 280(21), 20310-5.

Poon, R. T., K. K. Chung, S. T. Cheung, C. P. Lau, S. W. Tong, K. L. Leung, W. C. Yu, G. P. Tuszynski & S. T. Fan, 2004. Clinical significance of thrombospondin 1 expression in hepatocellular carcinoma. *Clin Cancer Res,* 10(12 Pt 1), 4150-7.

Ricciardelli, C. & R. J. Rodgers, 2006. Extracellular matrix of ovarian tumors. *Semin Reprod Med,* 24(4), 270-82.

Schmidt, W. M., M. Kalipciyan, E. Dornstauder, B. Rizovski, R. Sedivy, G. G. Steger, M. W. Muller & R. M. Mader, 2003. Gene expression profiling of colon cancer reveals a broad molecular repertoire in 5-fluorouracil resistance. *Int J Clin Pharmacol Ther,* 41(12), 624-5.

Schmidt, W. M., M. Kalipciyan, E. Dornstauder, B. Rizovski, G. G. Steger, R. Sedivy, M. W. Mueller & R. M. Mader, 2004. Dissecting progressive stages of 5-fluorouracil resistance in vitro using RNA expression profiling. *Int J Cancer,* 112(2), 200-12.

Shain, K. H. & W. S. Dalton, 2001. Cell adhesion is a key determinant in de novo multidrug resistance (MDR): new targets for the prevention of acquired MDR. *Mol Cancer Ther,* 1(1), 69-78.

Stefkova, J., R. Poledne & J. A. Hubacek, 2004. ATP-binding cassette (ABC) transporters in human metabolism and diseases. *Physiol Res,* 53(3), 235-43.

Sturn, A., J. Quackenbush & Z. Trajanoski, 2002. Genesis: cluster analysis of microarray data. *Bioinformatics,* 18(1), 207-8.

Szakacs, G., J. K. Paterson, J. A. Ludwig, C. Booth-Genthe & M. M. Gottesman, 2006. Targeting multidrug resistance in cancer. *Nat Rev Drug Discov,* 5(3), 219-34.

Toole, B. P., 2004. Hyaluronan: from extracellular glue to pericellular cue. *Nat Rev Cancer,* 4(7), 528-39.

Tringler, B., C. Grimm, G. Sliutz, S. Leodolter, P. Speiser, A. Reinthaller & L. A. Hefler, 2005. Immunohistochemical expression of thrombospondin-1 in invasive vulvar squamous cell carcinoma. *Gynecol Oncol,* 99(1), 80-3.

Uitto, J., 2005. The gene family of ABC transporters--novel mutations, new phenotypes. *Trends Mol Med,* 11(8), 341-3.

Wu, Y. J., D. P. La Pierre, J. Wu, A. J. Yee & B. B. Yang, 2005. The interaction of versican with its binding partners. *Cell Res,* 15(7), 483-94.

Yasui, K., S. Mihara, C. Zhao, H. Okamoto, F. Saito-Ohara, A. Tomida, T. Funato, A. Yokomizo, S. Naito, I. Imoto, T. Tsuruo & J. Inazawa, 2004. Alteration in copy numbers of genes as a mechanism for acquired drug resistance. *Cancer Res,* 64(4), 1403-10.

Zheng, P. S., J. Wen, L. C. Ang, W. Sheng, A. Viloria-Petit, Y. Wang, Y. Wu, R. S. Kerbel & B. B. Yang, 2004. Versican/PG-M G3 domain promotes tumor growth and angiogenesis. *Faseb J,* 18(6), 754-6.

7. GENOME SIGNATURES ASSOCIATED WITH POST-TREATMENT RECURRENCE IN EARLY STAGE LUNG CANCER[17]

7.1. Introduction

A quarter of worldwide cancer deaths result from lung cancer, with non-small cell lung cancer accounting for approximately 80% of cases (Canada 2007; Jemal et al. 2008; Travis 2002). More than two thirds of NSCLC patients present with locally advanced disease or distant metastases, precluding tumour resection as a curative treatment (Ries et al. 2008). Even half of resectable early stage tumours will recur, with adjuvant chemotherapy only slightly improving outcome (Ries et al. 2008). There is a clear need to be able to parse early stage patients based on their likelihood of recurrence following standard treatments; tools subclassifying lung cancer patients in this way would help direct application of chemotherapy only where it would have an effect and would help identify patients best-suited to clinical trials with novel regimens.

For lung cancer, previous attempts to identify molecular markers predicting disease recurrence or outcome have had mixed success. Single gene markers have been most effective where they have been the direct target of small molecular inhibitors or antibody-based treatments, with *EGFR* and various tyrosine kinase inhibitors representing the most prominent example (though there have been even ambiguous results in this relatively straightforward paradigm) (Karamouzis et al. 2007). Complex activation of nucleotide excision repair factors has been associated with different outcomes for more conventional treatments: activation of DNA repair factors such as *ERCC1* and *RRM1* is associated with a more favourable outcome in patients treated by surgery alone, but a poorer outcome where adjuvant platinum-based chemotherapy is also used (Gazdar 2007).

[17] A version of this chapter is being prepared for submission as a research manuscript: Timon P. H. Buys, Julia M. Chae, Chang Qi Zhu, William W. Lockwood, Bradley P. Coe, Raj Chari, Luc Girard, John Yee, John English, Sukhinder Atkar-Khattra, Dorothy Hwang, May Zhang, Cindy Dong, Heather Saprunoff, Jennifer Y. Kennett, Raymond Ng, Adi F. Gazdar, John D. Minna, Ming-Sound Tsao, Victor Ling, Calum MacAulay, Stephen Lam, Wan L. Lam. (2009) "Genomic signatures associated with post-treatment recurrence in early stage lung cancer," *in preparation*.

Previous data indicate that pro-metastatic molecular alterations may exist before distant metastases occur (Ramaswamy et al. 2003). In addition, we have uncovered genome alterations in pre-invasive oral lesions that may be associated with invasive disease (unpublished results). Genome alterations mediating response to chemotherapy may similarly exist in pre-treatment cancer tissue. If such changes are tied to other essential tumourigenic processes, they will likely occur at high incidence in a tumour cell population and be readily detected by analysis of that entire population.

We have evaluated global genomic alterations for a panel of early stage lung tumours treated by surgical resection and adjuvant chemotherapy with a *cis*-platinum/ vinorelbine doublet. Cases were obtained from multiple hospitals and had a minimum of two years follow-up. We investigated behaviour of individual gene candidates and global dysregulation of established molecular signaling pathways. Recent data suggest that specific genome alterations are associated with different lung cancer subtypes, a finding which has broad implications considering uniform approaches to NSCLC treatment (Kwei et al. 2008; Petersen et al. 1997; Weir et al. 2007). Because of this, we undertook independent analyses of genome alterations within histological groups in addition to a combined NSCLC analysis. Ultimately, alterations at multiple chromosomal regions were associated with tumour recurrence following treatment, with candidate loci spanning either potent oncogenes or multiple genes contributing to a given molecular signaling cascade.

7.2. Materials and methods

7.2.1. Tissue collection and DNA extraction

Three hundred seventy-four stage IA-IIIA NSCLC cases were obtained from the Ontario Cancer Institute and the British Columbia Cancer Agency. Tissue samples were obtained either from biopsy specimens (formalin-fixed, paraffin embedded material) or from freshly resected tumours. To limit the impact of tissue heterogeneity on DNA analysis, lung pathologist reviewed all cases and guided microdissection so that a minimum of 70% tumour cells contributed to each DNA extraction (these stringent criteria necessitated laser capture microdissection for a subset of cases). DNA was

isolated from each case by a standard phenol:chloroform protocol (Sambrook et al. 1989). Genome profiles were successfully generated for 198 of these tumours. Some of these tumours were from patients treated with a combination of surgery and adjuvant chemotherapy (*cis*-platinum and vinorelbine, n = 81) while others came from patients treated by surgery alone (n = 117). All cases were associated with detailed demographic data (e.g. age, sex, smoking history, etc.) (Table 7.1, 7.2). These cases all had >2 years follow-up and were evaluated for post-treatment recurrence. Recurrent disease was documented for approximately a third of cases (40/117 cases treated by surgery alone and 27/81 cases receiving surgery and chemo).

7.2.2. Genome profiling and data normalization

An array comparative genomic hybridization (aCGH) platform comprised of 26,363 tiled BAC clones was used to analyze tumour genomes (Watson et al. 2007). In brief, 200 ng of patient DNA and 200 ng of a standard individual male genomic reference DNA sample were labeled with cyanine-3 and cyanine-5 dCTP dyes respectively. Labeled probes were next combined, with unincorporated nucleotides removed by size exclusion (YM-30 Microcon centrifugation Tube, Millipore, Etobicoke, ON). Following addition of 100 µg C_0t-1 DNA (Invitrogen, Carlsbad, CA), this mixture was precipitated and then resuspended in a 45 µl cocktail including DIG Easy hybridization solution (Roche, Mississaga, ON), yeast tRNA (Calbiochem, Mississaga, ON), and sheared herring sperm DNA (Sigma-Aldrich, Oakville, ON). Steps for probe denaturing and blocking were then perfromed at 85°C and 45°C for 10 minutes and for 1 hour respectively. Following this, the probe mixture was applied to the surface of the CGH array, coverslips were put in place, and arrays were incubated for 36-40 hours at 45°C. After incubation came five agitating washes in 0.1X saline sodium citrate, 0.1% sodium dodecyl sulfate at 45°C (each wash ~5 min). This was followed by rinses with 0.1X SSC and drying by centrifugation, then scanning of microarray images by a CCD-based camera system. Microarray images were analyzed using SoftWoRx Tracker Spot Analysis software (Applied Precision, Issaquah, WA). Experimental biases due to the array platform and DNA samples of varying quality were normalized as previously described (Chi et al. 2008; Khojasteh et al. 2005). Log_2 signal intensity ratios for each

clone on the array were visualized against their known chromosomal position using SeeGH software (Chi et al. 2008). Clones were filtered from analysis when their signal-to-noise ratios were <3 or standard deviations between replicate spots were >0.1.

7.2.1. Analysis of genomic data

The aCGH-Smooth segmentation algorithm delineated segmental alterations for each case (Jong et al. 2004). Default parameters were used on all settings except for the following changes: lambda value = 6.75, maximum number of breakpoints in initial pool = 100. Analysis was restricted to autosomes to avoid bias due to use of an individual male DNA reference sample. Each clone was called as gained, lost, or unchanged. These values were then used to determine frequency of gain and loss across a given sample panel (e.g. recurring NSCLC cases). Fisher's Exact Test was then used to identify genomic tracts with significantly different activation ($p < 0.05$ cut-off).

7.2.2. Gene expression analysis

Gene expression profiles for each tumour specimen were generated using a custom oligonucleotide microarray (Agilent). For each case, 5 µg of RNA from a given tumour sample and a universal reference RNA sample were reverse transcribed to cRNA, labeled with the fluorescent dCTP dyes cyanine-3 and cyanine-5 (replicate flip-fluor experiments were performed), and competitively hybridized to the microarray surface. Data were normalized using Rosetta Resolver software and $\log_{10}$ transformed (Vardhanabhuti et al. 2006). Significant differences in expression between groups of tumours were calculated using a Mann-Whitney *U* test ($p < 0.05$).

7.2.3. Parallel analysis of DNA alterations and pathway disruption

The approaches applied to identify genes associated with post-treatment recurrence and chemoresistance are shown in Figure 7.1 and 7.2. Functional annotation and differential pathway activation for these genes was determined using Ingenuity Pathway Analysis software (Ingenuity). Fisher's Exact Test was used to determine whether disruption of a given fraction of genes in a specific pathway was significant ($p < 0.05$).

7.2.4. Validation in NSCLC cell lines

Forty-nine lung cancer cell lines were used to further investigate genome alteration findings from clinical samples. These lines were a generous gift from Dr. John Minna (University of Texas, Southwestern), whose lab determined *cis*-platinum sensitivity and vinorelbine sensitivity for each cell line. Briefly, seeded cells were treated with drug and their viability was assessed by the MTT (3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide) assay. Replicate experiments using varying doses of each drug were used to define IC_{50} values and these values were used to define each line as "sensitive", "intermediate", or "resistant" to each drug. Genomic data were generated and analyzed as above. Gene expression profiles were generated using the Affymetrix U133 Plus 2.0 array platform (data were normalized using Microarray Suite 5.0). Differences in gene expression between drug response groups were calculated by Mann-Whitney *U* test ($p < 0.05$). Candidate genes were only included where gene expression changes matched the direction that would be expected based on DNA copy number changes (i.e. "gained" genes had to be over-expressed and "lost" genes had to be under-expressed).

7.3. Results

7.3.1. Identification of DNA alterations associated with recurrence following adjuvant chemotherapy

We sought to identify genomic alterations associated with tumour recurrence in patients who received treatment with adjuvant *cis*-platinum/ vinorelbine. To that end, we analyzed 198 resectable lung cancer cases by tiling-path array CGH (a representative genomic profile is shown in Figure 7.3) (Watson et al. 2007). One hundred seventeen of these cases were treated by surgery alone (Table 7.1) and the remaining 81 received both surgery and adjuvant chemotherapy in the form of a *cis*-platinum/vinorelbine doublet (Table 7.2). Sample accrual and genome profiling approaches are described in the Materials and Methods. The experimental framework used to delineate DNA changes associated with recurrence following chemotherapy is depicted in Figure 7.1.

Initially, we defined genome alterations associated with post-adjuvant chemotherapy recurrence in a combined panel NSCLC tumours. For tumours from patients treated with surgery and adjuvant chemotherapy, a comparison of the genomic alterations in recurring and non-recurring cases showed differential alteration in 7% of the genome (Fisher's Exact Test, $p < 0.05$) (Figure 7.4). Lung tumour histological subgroups, which have previously been demonstrated to harbour distinct DNA changes, were represented at different frequencies within in each recurrence group of this panel (e.g. squamous tumours represented 42% and 22% of recurring cases and non-recurring cases respectively). To adjust for this imbalance, we compared genomic alterations for squamous and non-squamous tumours in this same panel of tumours. Regions found to be significantly altered by both analyses – recurrence and histological subtype – were excluded from further analysis. This same analytical approach was applied to the panel of 117 lung tumours that were treated by surgery alone (Supplementary Figure 7.S1.), leaving us with recurrence-associated alterations in each lung tumour panel.[18] Finally, to hone in on alterations associated with recurrence after adjuvant chemotherapy, we subtracted changes identified in the panel of tumours treated by surgery alone from those identified in the panel of tumours treated by surgery and a *cis*-platinum/ vinorelbine doublet. This resulted in a shortlist of 242 genes that were significantly associated with chemoresponse, ~94% of which showed relative activation in recurring cases (i.e. higher frequency of gain in recurring cases and/or higher frequency of loss in non-recurring cases) (Supplementary Table 7.S1). Pathway analysis software for these data indicated that the BMP signaling cascade was strongly disrupted, with various components deleted in non-recurring tumours (Figure 7.5).

Given evidence of different underlying genome changes in histological subtypes, we also undertook separate analyses for squamous and non-squamous tumours. The analytical approaches for genomic data were undertaken as above, the only change being that exclusion of genome alterations associated with histological subtypes was excluded (Figure 7.1). For each subtype, genome alterations significantly associated

[18] All Supplementary Figures and Tables for Chapter 7 are included in Appendix 1.

with recurrence after surgery and adjuvant chemotherapy were identified (Figure 7.6). No overlap was observed between these two groups. (Results for subtype-specific analysis for cases treated by surgery alone are shown in Supplementary Figure 7.S2.) The most prominent alteration associated with recurrence in non-squamous tumours was gain of chromosome arm 7p, which was more commonly gained in recurring cases (31 non-recurring cases were compared to 21 recurring cases for this analysis). Several other regions also harboured significantly different regions of alteration (see Supplementary Table 7.S2). Regions of alteration significantly associated with recurrence were also identified for squamous tumours (23 non-recurring versus 6 recurring), including chromosome arm 4q gain and chromosome arm 8p loss, which were both associated with post-treatment recurrence (see Supplementary Table 7.S3). Combined pathway activation analysis for genes defined as significantly altered in either histological group revealed significant losses for components of the BMP signaling pathway. This was similar to what was observed when all NSCLC cases were analyzed together, though different pathway components were dysregulated (Figure 7.7).

7.3.2. Analysis of matched gene expression data from lung tumours

Genes identified by the above genomic analyses were further investigated in mRNA expression profiles which were available for 9/81 NSCLC cases. Five of these cases did not recur following treatment with surgery and chemotherapy while four did recur. Of the genes significantly associated with recurrence in NSCLC by genomic analysis, only eight were found to have significant expression changes in the same direction as the DNA alterations (Table 7.3). We repeated this analysis for genes significantly altered at the genomic level in non-squamous tumours (only one expression profile was from a squamous tumour, therefore analysis could not be undertaken for that subtype). In this comparison of four recurring and four non-recurring cases, only two genes found to be significantly different at the genomic level were also found to be significantly altered at the gene expression level: *KIAA0895* and *EGFR*.

7.3.3. Analysis of genomic dysregulation in cell models

Having identified gene alterations associated with recurrence following adjuvant chemotherapy, we next sought to associate these changes with resistance to the specific components of the adjuvant chemotherapy (the experimental approach for this is outlined in Figure 7.2). We examined the status of genomic alterations identified in clinical lung tumours in lung cancer cell lines with known response to drug (Supplementary Tables 7.S1, 7.S2, 7.S3).[19] IC_{50} values in a panel of 49 NSCLC tumour lines were known for *cis*-platinum (CDDP) and vinorelbine (Table 7.4, Materials and Methods). For CDDP, six lines were defined as sensitive to treatment while 11 were defined as resistant. For vinorelbine, 42 were defined as sensitive and seven were defined as resistant. Genome profiles for each of these lines were generated and analyzed by the same approaches applied to lung tumours. Results are shown in Supplementary Figure 7.S3. No genes were found to be significantly associated with both post-treatment recurrence in tumours and vinorelbine resistance in cell lines. With respect to CDDP, only one gene was shown to exhibit the same alteration status in clinical specimens and CDDP-resistant cell lines; *KIAA1303*, also known as *Regulatory-associated protein of mTOR* (*Raptor*), was found to be gained in both groups.

This analytical approach in cell lines was then repeated to validate our findings in non-squamous tumours.[20] With respect to vinorelbine response in non-squamous cell lines, multiple regions of genomic alteration were identified as significantly difference following the comparison of 37 sensitive lines to 7 resistant lines (Supplementary Figure 7.S4). However none of these regions intersected with regions described as significantly altered with response in clinical specimens. Analysis for CDDP-resistant non-squamous lines was not repeated as all cases represented in the analysis above were already of this histological grouping (Supplementary Figure 7.S3A). Sixty-one genes associated with recurrence following surgery and chemotherapy were also associated

[19] Because so few tumours had matched mRNA expression microarray data, analysis was not restricted to candidate genes identified by both genomic and gene expression analysis of tumours.
[20] Analysis of squamous cell lines was precluded since none of the five lines with this histology were classified as resistant to either drug tested.

with CDDP-resistance in non-squamous cell lines (Table 7.5). All of these genes were located on chromosome 7 and were gained where recurrence/resistance occurred.

7.3.4. Analysis of candidate gene expression in NSCLC cell lines

Matched gene expression profiles were available for most analyzed lung cancer cell lines. With respect to the combined analysis of NSCLC cases, *Raptor* was not found to have significantly different expression between CDDP-resistant and CDDP-sensitive cells. With respect to the analysis of non-squamous tumours, only ten of the chromosome 7 genes identified above were found to have similar behaviour at the gene expression level (Figure 7.8). *EGFR* was the only gene found to be significantly altered in terms of DNA copy number and gene expression in both clinical samples and cell lines, its activation apparent in non-squamous tumours resistant to CDDP. No genes were found to behave this way in the combined analysis of all NSCLC samples and data were insufficient for this kind of analysis in squamous tumours.

7.4. Discussion

Gene candidates associated with recurrence after adjuvant chemotherapy were identified in a panel of 198 resected lung tumours. This was accomplished by excluding genomic alterations associated with specific tumour histologies and post-surgical recurrence (Table 7.1 and 7.2, Figure 7.1). Imbalances in the incidence of histological groups within different recurrence groups were also addressed by independent analyses for chemoresponse genes in squamous tumours and non-squamous tumours within this panel of NSCLC cases (Figure 7.6).

7.4.1. Molecular candidates associated with post-adjuvant chemotherapy recurrence by combined analysis of NSCLC tumours

Analysis of molecular pathway activation in NSCLC tumours indicated that the BMP signaling cascade was significantly down-regulated in non-recurring cases and/or activated in recurring cases (Figure 7.5). (See also *SMAD1* gain, Figure 7.3.) BMP pathway signaling has previously been implicated in tumour suppression for a variety of cancer types, though its broad array of downstream targets suggest it can be involved in

a variety of cellular phenotypes (Hardwick et al. 2008; Kim & Kim 2006). While BMP signaling has previously been demonstrated to mediate a variety of key oncogenic processes in a variety of cancer types, it has not previously been implicated in response to CDDP or vinorelbine (Hardwick et al. 2008; Gelse et al. 2008; Langenfeld et al. 2006). Although relationships between some of the identified genes can be antagonistic, the complex dysregulation we observed may represent a means for focusing signals to specific downstream targets (Kitisin et al. 2007). Pathway analysis specifically showed that six BMP pathway genes were differentially altered between recurrence groups; approximately half the tumours that did not recur had deletions spanning at least one of these genes. Whether this suggests that retention of BMP signaling is necessary for recurrence or loss of BMP signaling fosters chemosensitivity remains to be determined.

Individual genes identified in this combined NSCLC analysis may also play a role in mediating post-treatment recurrence. *XRCC4*, a component of the non-homologous end joining (NHEJ) DNA repair pathway, was lost in 36% of tumours that did not recur following treatment but retained or gained in a majority of cases that did recur. Polymorphisms at *XRCC1*, another NHEJ component, have recently been implicated in response to platinum-based treatment (Sun et al. 2009). Given that *XRCC4* and *XRCC1* both play a critical role in repairing double stranded DNA breaks, this retention of *XRCC4* in recurring cases could be significant for countering the effectives of DNA-damaging CDDP (Audebert et al. 2004; Burma et al. 2006).

Integrative analysis of genomic and gene expression data was possible for nine NSCLC tumours, with eight genes found to be significantly altered in the same direction (e.g. gained and over-expressed) (Table 7.3). Only one of these genes – *HABP2* – was previously associated with lung cancer, though it had not been reported as associated with prognosis or response to chemotherapy (Chong et al. 2006; Wang et al. 2002). While neither BMP pathway members nor *XRCC4* were identified by this integrative approach, this absence may have occurred because only a small fraction of cases were represented by expression microarray profiles; it is quite possible that these genes

would be detected as significantly associated with post-treatment recurrence if additional matched expression data were available.

7.4.2. DNA copy number changes associated with recurrence following treatment with adjuvant *cis*-platinum/ vinorelbine in lung tumour histological subtypes

Different molecular alterations may drive the emergence of different NSCLC histological subtypes. Subtype-specific gene changes – such as gain of chromosome 3q in lung squamous cell carcinomas – suggest that NSCLC may contain multiple biologically distinct tumour types, each warranting independent analysis and treatment approaches (Kwei et al. 2008; Petersen et al. 1997; Weir et al. 2007). To that end, we identified genome alterations associated with recurrence after adjuvant chemotherapy separately for squamous and non-squamous tumours (as defined above). Distinct, non-overlapping DNA changes were found associated with each subgroup (Figure 7.6). Some BMP signaling pathway genes were disrupted in these subtypes, though these were different genes than had been observed for the combined NSCLC analysis (Figure 7.7). Disruption of different components of the same receptor complex in non-recurring tumours (*BMPR1A* in non-squamous tumours and *BMPR1B* in squamous tumours) could be a product of co-activation with neighbouring genes, with these broader genome changes specific to each tumour subtype. Taken together, both the combined NSCLC analysis and the independent histological subtype analyses suggest that low level disruption at multiple loci from this pathway may play a role in chemosensitivity.

Nearly 70% of the genes associated with post-treatment recurrence in non-squamous tumours were found on chromosome 7p, which was gained in >30% of recurring cases (Supplementary Table 7.S2). Retention/gain of *XRCC4* was found to be significantly associated with recurrence in squamous tumours, as it was in the combined NSCLC analysis. Similarly, *SFRS1* – a splicing factor previously described as a proto-oncogene in multiple cancer types – was found to have relative activation in recurring samples following both the combined NSCLC and the separate non-squamous tumour analyses (Karni et al. 2008). These results suggest that some of the alterations detected by

combined NSCLC analysis may have been prominent differences within histological subgroups. Detection of compelling recurrence-associated genes for each tumour subtype supports the idea that histological groups represent biologically distinct disease classes.

Gene expression data were combined with genomic data for integrative analysis of non-squamous tumours, as was done with the combined NSCLC analysis (this analysis was precluded for squamous tumours as only one expression microarray profile was available). Comparison of expression profiles for four recurring tumours against four non-recurring tumours revealed only two genes to be significantly associated with genomic and gene expression changes. One gene encoded the hypothetical protein *KIAA0895*, located at 7p14.1. The function of this gene is unknown and it has not previously been implicated in lung cancer. The other gene was *EGFR* which has previously been described as amplified in multiple NSCLC subtypes and is described in greater detail below. When additional expression data for these cases becomes available, this analysis will be repeated to confirm our findings.

7.4.3. Investigation of chemoresistance genes in lung cancer cell models

Vinorelbine and CDDP sensitivity was known for a panel of 49 lung tumor cell lines with matched genomic and gene expression data (Table 7.4). Drug sensitive and resistant lines were compared to identify gene changes significantly associated with chemoresistance. Separate analyses allowed association of specific gene changes with response to each drug. Genes identified *in vitro* were then compared to genes we associated with post-treatment recurrence in tumours (Figure 7.2). *Raptor*, which encodes a component of the mTORC1 complex and plays a role in mTOR-mediated regulation of mRNA translation, was the only gene found to be altered at the genomic level in both a combined analysis of NSCLC and an analysis of CDDP-resistant NSCLC cell lines (Sabatini 2006). No genes associated with tumour recurrence in clinical samples were similarly altered in vinorelbine-resistant lung cancer cells. Genes uncovered by the combined analysis of all NSCLC tumours may not have been detected in cell lines because of differences in the histological make-up of the cell line

panel, which only had five squamous tumour lines (none of which were associated with resistance to either CDDP or vinorelbine).

In regard to previously described genes associated with recurrence following application of adjuvant platinum-based chemotherapy, *ERCC1*, *RRM1*, and *BRCA1* copy number alterations were not found to be significantly associated with post-treatment recurrence in combined analysis of NSCLC or histology subtype-specific analysis (Rosell et al. 2006). Neither were these genes found to be associated with chemoresistance in cell lines (which are discussed in greater detail below).

7.4.4. Implication of recurrence-associated genes from non-squamous tumours in chemoresistance in lung cancer cell lines

Genomic data for cell lines only overlapped with data from recurring lung tumours in the case of CDDP-resistant non-squamous tumours (explained in Results). It has previously been suggested that aneuploidy may drive various cancer phenotypes, including chemoresistance (Duesberg 2005; Duesberg et al. 2007). Certainly there is evidence in lung tumours that there may be several oncogenes on chromosome 7 (Campbell et al. 2008; Garnis et al. 2006). All 61 genes uncovered by integration of genomic data from recurring non-squamous lung tumours and CDDP-resistant cancer cell lines were located on chromosome 7p (Table 7.5). Of the ten genes on this list that also exhibited increased expression in CDDP-resistant lines (Figure 7.8), many of these had previously been implicated in cancer (though only *HUS1* and *EGFR* have previously been implicated in chemoresistance) (Di Renzo et al. 2006; Epel et al. 2008; Floyd et al. 1998; Hoque et al. 2006; Katz et al. 2007). *HUS1* has been shown to complex with Rad1 and Rad9 to repair DNA damage (Volkmer & Karnitz 1999). *HUS1* has also previously been shown to mediate sensitivity to CDDP in lung cancer cells and has also been implicated in regulating radiosensitivity (Brandt et al. 2006; Kinzel et al. 2002). Its status and association with outcome in clinical lung cancer has not previously been evaluated and warrants further analysis.

It is striking that only *EGFR* exhibited the same genomic and transcriptional activation in both clinical specimens and tumour cell lines. Previous studies have implicated *EGFR*

activation in CDDP resistance in neuroblastoma, head and neck, and breast cancer cell lines (Eckstein et al. 2008; Michaelis et al. 2008; Sok et al. 2006). Additionally, *EGFR* shares downstream targets with the BMP pathway and has been demonstrated to signal via mTOR (Faivre et al. 2006; Jiang & Liu 2008; Sabatini 2006). While previous analysis of *EGFR* has suggested that copy number change is not associated with progression or survival following platinum-based or non-platinum based chemotherapy, the sample sizes used for these analyses were quite small and in many instances was undertaken with advanced disease stages (Cappuzzo et al. 2007; Ceppi et al. 2006). Moreover these patients were of mixed histologies whereas we have restricted our analysis within tumour subtypes. Recently, analysis of lung tumour cell lines has indicated that previous CDDP exposure reduces sensitivity to erlotinib (Chin et al. 2008). These data in conjunction with our own suggest that tyrosine kinase inhibitors (TKIs) may prove useful as first-line adjuvant chemotherapy in early stage lung cancer, with *EGFR* copy number status used not just as a marker supporting application of TKIs but also as a means of excluding treatment with CDDP.

7.5. Conclusions

We have identified gene alterations in chemo-naïve lung cancer cells that are associated with recurrence following treatment with a *cis*-platinum/vinorelbine doublet. The fact that these DNA changes are detectable in a tumour cell population prior to treatment suggests that a majority of malignant cells could be drug resistant even before exposure to chemotherapy. This in turn suggests that gene changes driving chemoresistance may be intrinsic to other crucial tumourigenic processes; post-treatment recurrence may not be a product of clonal expansion for a small portion of resistant cells from the original tumour mass.

Our findings implicate BMP pathway signaling in recurrence following treatment with adjuvant *cis*-platinum and vinorelbine. These results also justify independent genomic analysis of lung tumour histological subtypes and reinforce the idea that different NSCLC subgroups represent biologically distinct forms of disease. Also, our data indicate that *EGFR* gene copy number may be useful for guiding treatment decisions

outside of the context of TKI therapy. Analysis in an independent panel of non-squamous lung tumours is needed to confirm the association between *EGFR* gain/amplification and recurrence following platinum-based adjuvant therapy. Future work will also evaluate the utility of increased *EGFR* copy number as a CDDP chemoresponse marker relative to existing markers such as *ERCC1*, *RRM1*, and *BRCA1* expression. Our data may also be used as a rationale for investigating the use of TKIs in an adjuvant context for resectable lung cancer, particularly where molecular markers indicate that platinum-based modalities would be ineffective.

Table 7.1 – Demographic data for early stage NSCLC patients treated by surgery alone (n = 117).

Age*	Range Median	39 to 90 69
Stage*	I II III	75 28 9
Histology	Squamous Adenocarcinoma Large cell carcinoma Undifferentiated NSCLC	35 (30%) 71 (60%) 3 (3%) 8 (7%)
Recurrence status	Responsive disease Recurrence following treatment	77 (66%) 40 (34%)
Sex	Male Female	51 66
Smoking Status*	Current smoker Former smoker Never smoker	32 59 25

*Age, stage, and smoking status were unknown for a small number of cases.

Table 7.2 – Demographic data for early stage NSCLC patients treated by surgery and adjuvant chemotherapy (*cis*-platinum/vinorelbine doublet) (n = 81).

Age	Range Median	34 to 75 62
Stage	I II	41 (51%) 40 (49%)
Histology	Squamous Adenocarcinoma Large cell carcinoma Undifferentiated NSCLC	27 (33%) 42 (52%) 4 (5%) 8 (10%)
Recurrence status	No recurrence Recurrence following treatment	27 (33%) 54 (67%)
Sex	Male Female	43 (53%) 38 (47%)
Smoking Status*	Current smoker Former smoker Never smoker	28 (35%) 45 (56%) 7 (9%)

*Smoking status for one case was unknown.

Figure 7.1 – Experimental framework for genome analysis of lung tumours. We analyzed a panel of NSCLC tumours to identify genome alterations associated with recurrence following treatment with adjuvant chemotherapy (a *cis*-platinum/ vinorelbine doublet). Each case was treated either by resection plus adjuvant chemotherapy (Group 1) or by resection alone (Group 2). Cases were classified as recurring or non-recurring after treatment. Step 1: Within each group, genomic alterations in the recurring and non-recurring sub-groups were compared. This defined genomic alterations significantly associated with post-treatment recurrence. Step 2: Owing to sample size, histological subtypes were represented in different amounts within each group and within each recurrence sub-group (different tumour histological subtypes have previously been associated with specific chromosomal imbalances). To compensate for this potential bias, we defined genomic alterations associated with histology within each group, then filtered those regions from analysis. Step 3: To define alterations significantly associated with recurrence following adjuvant chemotherapy, we subtracted results for samples treated by surgery alone (Group 2) from results for samples treated by surgery plus adjuvant chemotherapy (Group 1). This same approach was also applied for analysis within lung tumour histological subgroups, with Step 2 omitted.

Figure 7.1

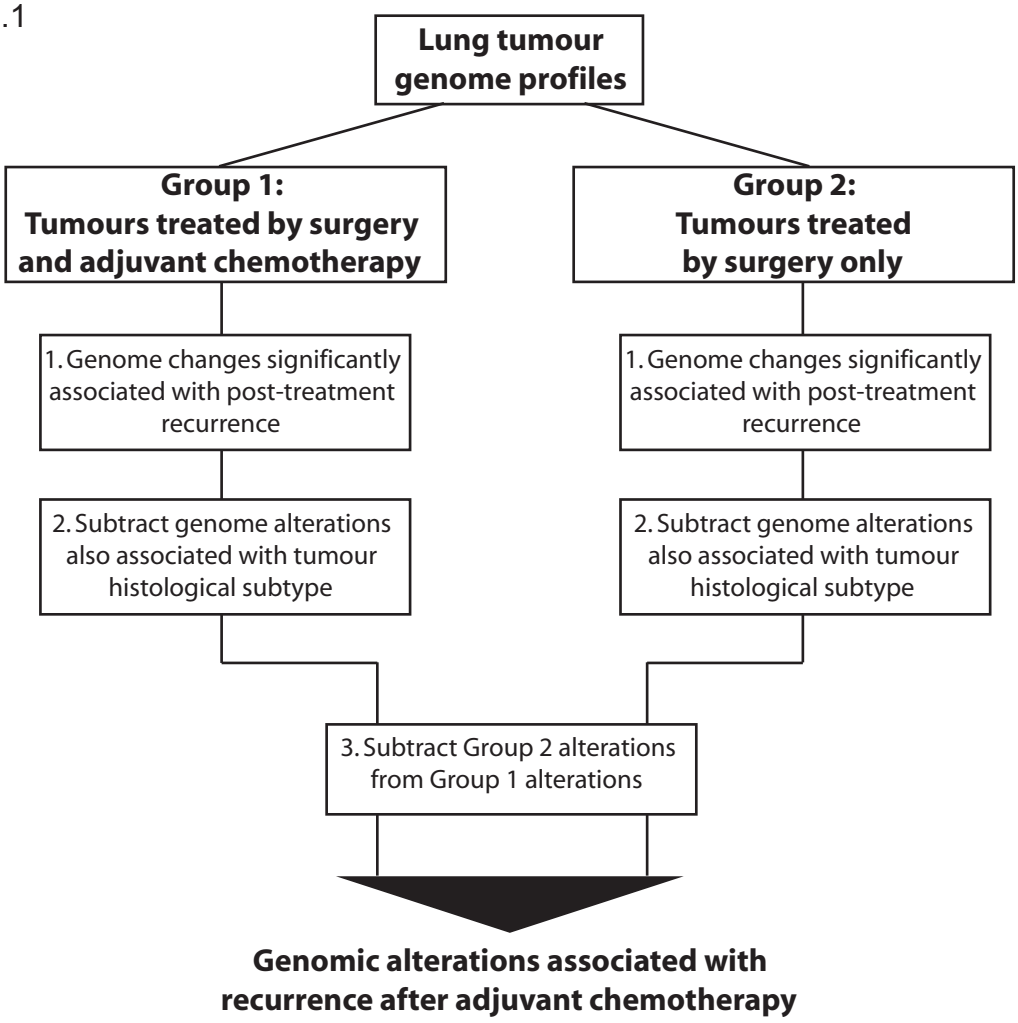

Lung tumour genome profiles
Group 1: Tumours treated by surgery and adjuvant chemotherapy
Group 2: Tumours treated by surgery only
1. Genome changes significantly associated with post-treatment recurrence
1. Genome changes significantly associated with post-treatment recurrence
2. Subtract genome alterations also associated with tumour histological subtype
2. Subtract genome alterations also associated with tumour histological subtype
3. Subtract Group 2 alterations from Group 1 alterations
Genomic alterations associated with recurrence after adjuvant chemotherapy

Figure 7.2

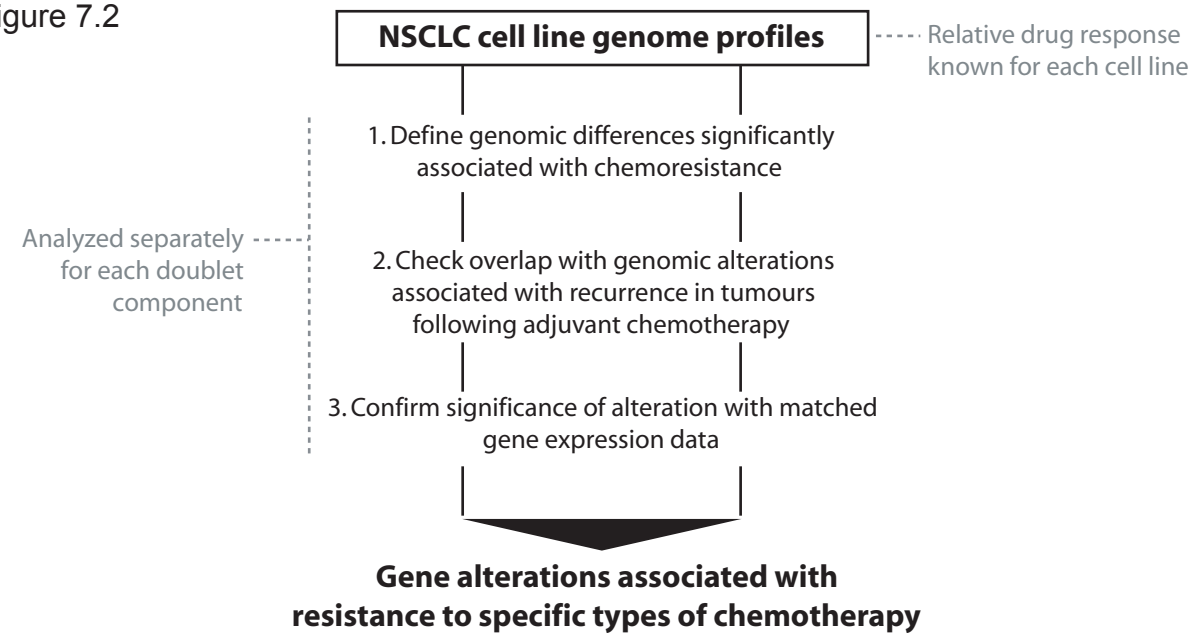


Figure 7.2 – Experimental framework for NSCLC cell lines. Separate IC_{50} values for *cis*-platinum and vinorelbine were known for a panel of NSCLC cell lines (see Materials and Methods). For each drug, cell lines were classified as having either a resistant, intermediate, or sensitive response to drug. Step 1: DNA alterations significantly associated with chemoresistance were defined by a comparison of genomic data for resistant and sensitive cell lines. Step 2: To establish clinical relevance, candidate chemoresistance genes from cell lines were then compared to genes already associated with recurrence in tumours following adjuvant chemotherapy. (The approaches used to define these latter alterations were already described.) Those genes appearing in both groups were analyzed further. Step 3: Matched expression data from cell lines were then evaluated to confirm activation of candidate chemoresistance genes. This approach was applied in both combined NSCLC analysis and analysis of separate tumour histology subtypes.

Figure 7.3 – Karyogram for a lung adenocarcinoma recurring after treatment with surgery and chemotherapy. Each BAC clone on the aCGH platform is represented as a blue spot at its known chromosomal position. Clones are plotted along the horizontal axis based on the $\log_2$ value of their signal intensity ratios from the array experiment. Vertical lines show $\log_2$ signal intensity ratios from -1 (green) to 1 (red), with copy number increases to the right of zero (purple line) and decreases to the left. A ratio of zero represents equivalent copy number between the hybridized samples. Where $\log_2 > 0$, it suggests increased DNA copy number in the tumour specimen and where $\log_2 < 0$ it suggests loss. Vertical pink bars span alterations encompassing *SMAD1* (4q31.22) and *EGFR* (7p11.2).

Figure 7.3

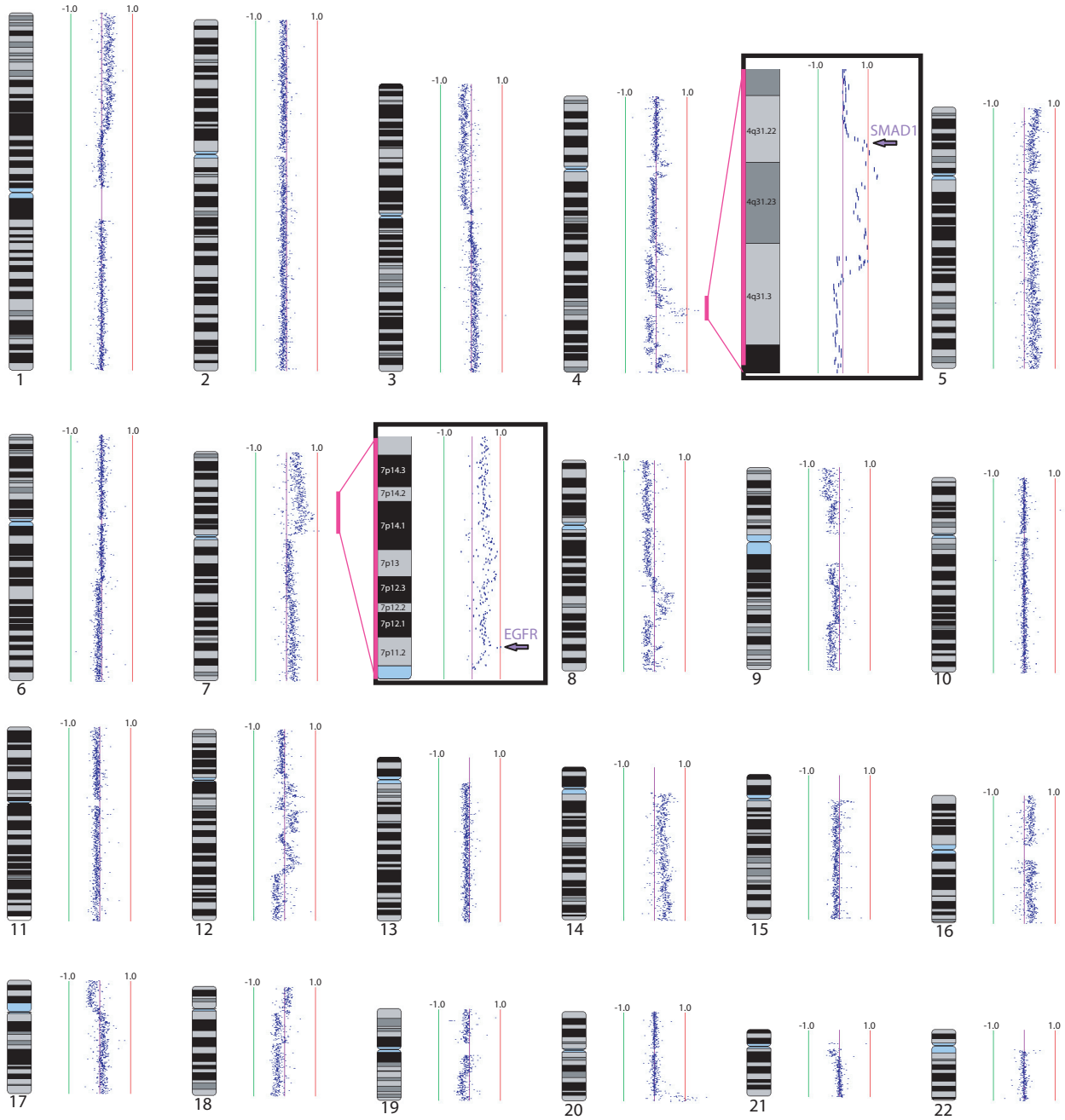

-1.0
1.0
4q31.22
4q31.23
4q31.3
SMAD1
7p14.3
7p14.2
7p14.1
7p13
7p12.3
7p12.2
7p12.1
7p11.2
EGFR
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22

Figure 7.4

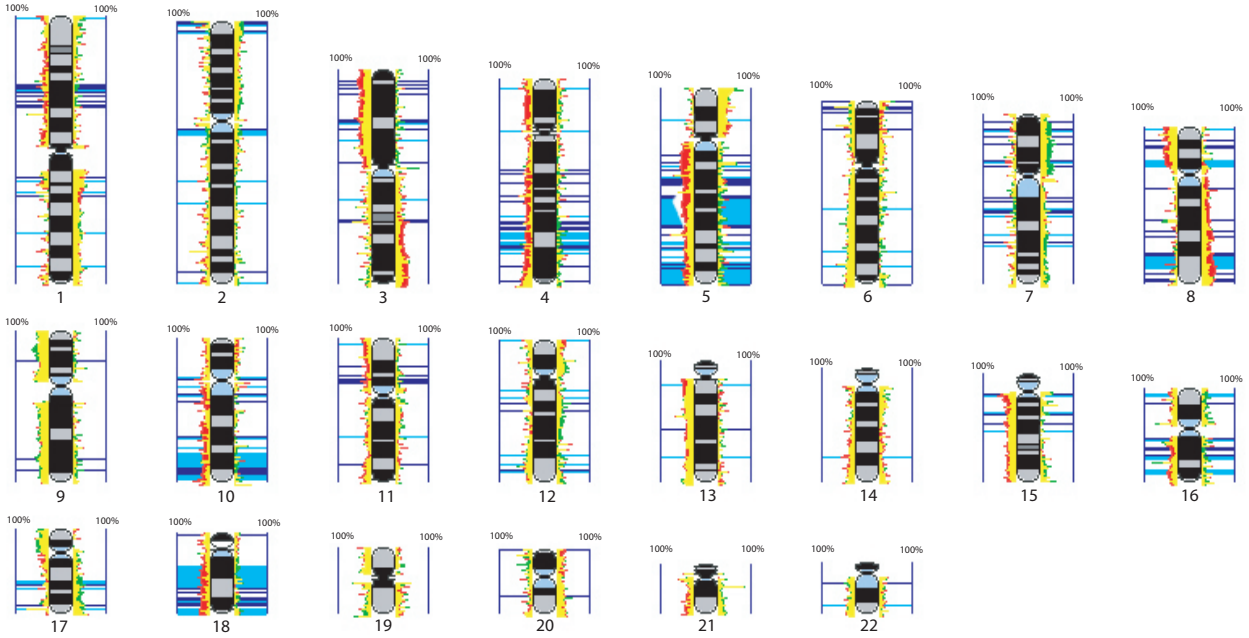


Figure 7.4 – Whole genome frequency of alteration for NSCLC patients treated with surgery and chemotherapy. Individual tumour genome profiles were generated for each case in a panel of stage IA-IIIA lung tumours (n = 81, Figure 7.3). Segmental DNA copy number alterations for each case were defined as described in Materials and Methods. Genome loci were altered to a maximum frequency of 100%. In this image, frequency of DNA gain is defined to the right of each chromosome, while frequency of loss is defined to the left. Red represents "non-recurring" specimens and green represents "recurring" (yellow indicates overlap between these two groups). The lighter blue horizontal lines in the karyogram span genome segments found to be significantly different between response groups and between histological groups (squamous versus non-squamous subtypes) (Fisher's Exact Test, $p < 0.05$). The dark blue horizontal lines span segments significantly associated with response groups alone.

Figure 7.5

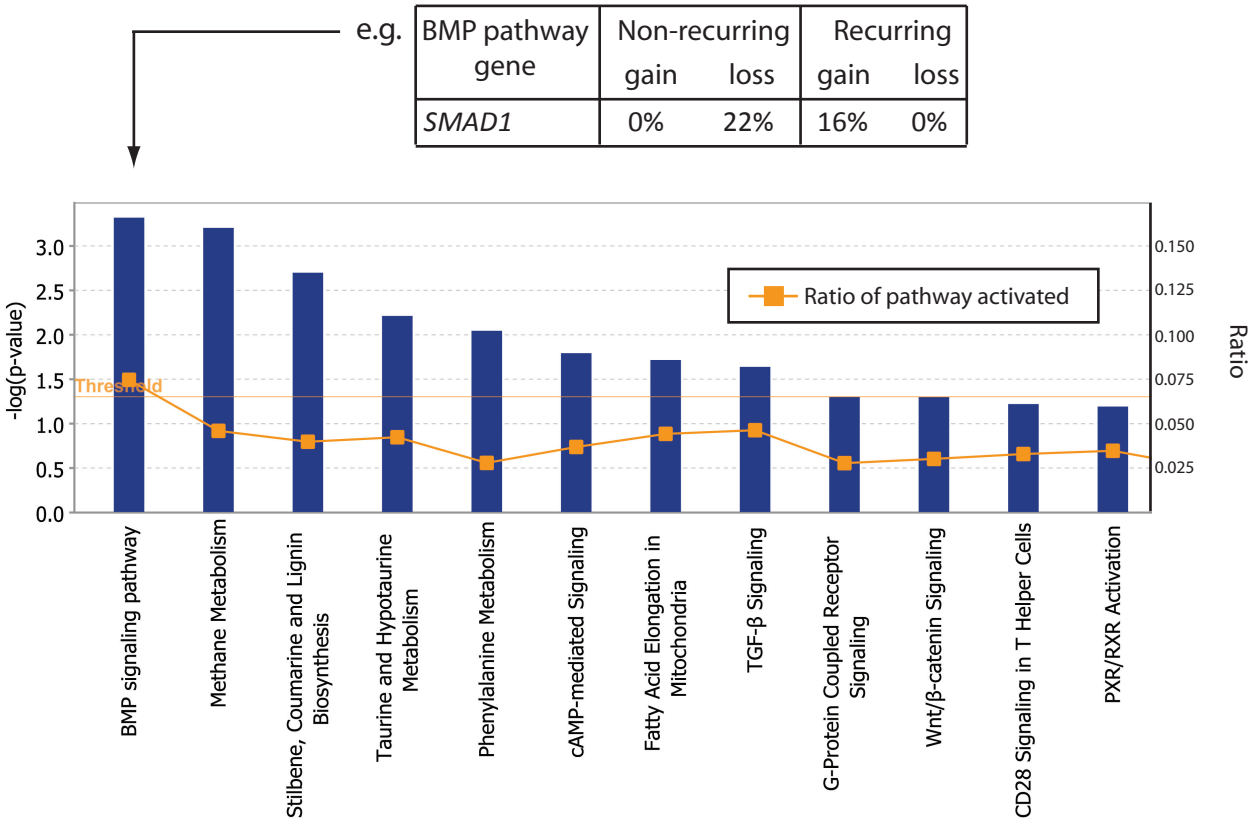
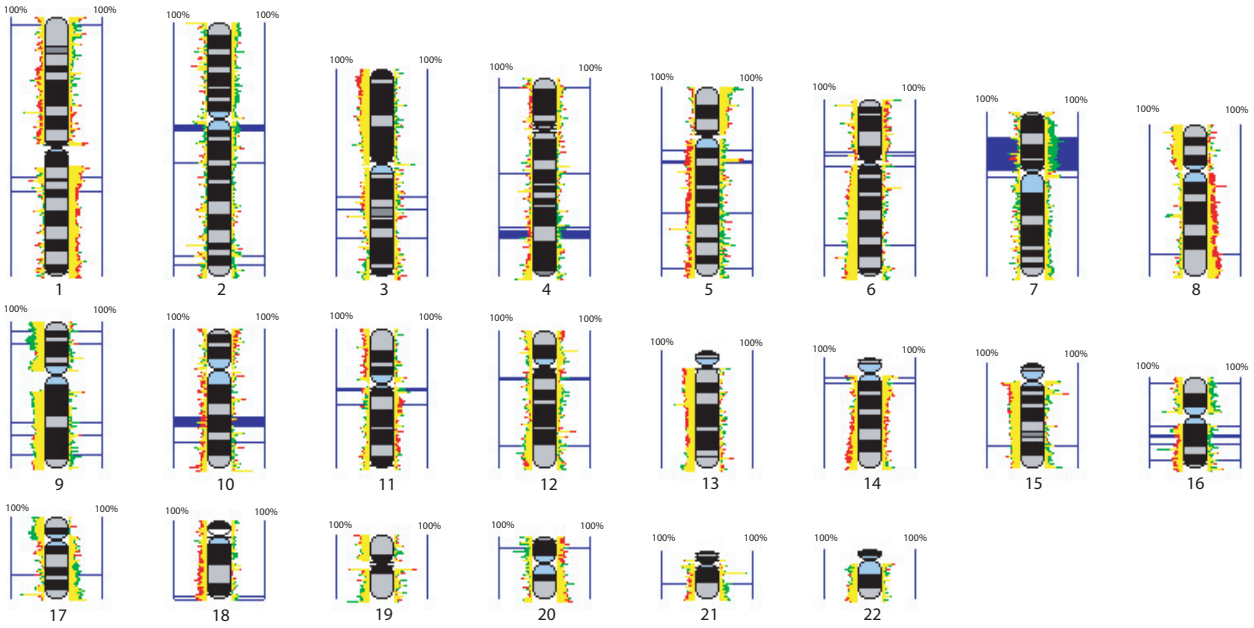


Figure 7.5 – Signaling pathways with relative activation in NSCLC cases recurring following treatment with chemotherapy. Genes associated with chemoresponse were assigned to their known signaling pathways using Ingenuity Pathway Assist software. The x-axis shows the pathway name. The right y-axis shows the ratio of genes affected in a given pathway (e.g. the ratio for the BMP signaling pathway was 0.075 [6/80]). The left y-axis shows positive values for log-transformed p-values calculated for each pathway using Fisher's Exact Test. Each blue bar represents the p-value for a given pathway. The threshold value was ~1.3 (orange horizontal line, -log[0.05]). Additional pathways not approaching the threshold value were not displayed. The table at the top shows an example of a gene contributing to the BMP signaling pathway that was significantly altered. Details about the frequency of alteration for this gene in different post-treatment recurrence groups is included.

Figure 7.6 – Whole genome frequency of alteration within NSCLC subtypes where patients were treated with surgery and chemotherapy. Both images are annotated as in Figure 7.4, with red representing “non-recurring”, green as “recurring”, and yellow indicating an overlap between these groups. Dark blue lines denote regions significantly different between recurrence groups (Fisher’s Exact Test, $p < 0.05$). A. Regions of differential genomic alteration within non-squamous tumours. Specifically, genomic differences were determined for 31 non-recurring versus 21 recurring tumours. Similar alterations identified by this analysis and by analysis of non-squamous tumours treated by surgery alone (51 non-recurring versus 31 recurring) were excluded from further analysis. B. Regions of differential genomic alteration within squamous tumours. As with non-squamous tumours, differences determined for patients receiving surgery alone (26 non-recurring versus 9 recurring) were subtracted from changes observed for patients treated by surgery and chemotherapy (23 non-recurring versus 6 recurring).

Figure 7.6

A.


100%
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22

B.

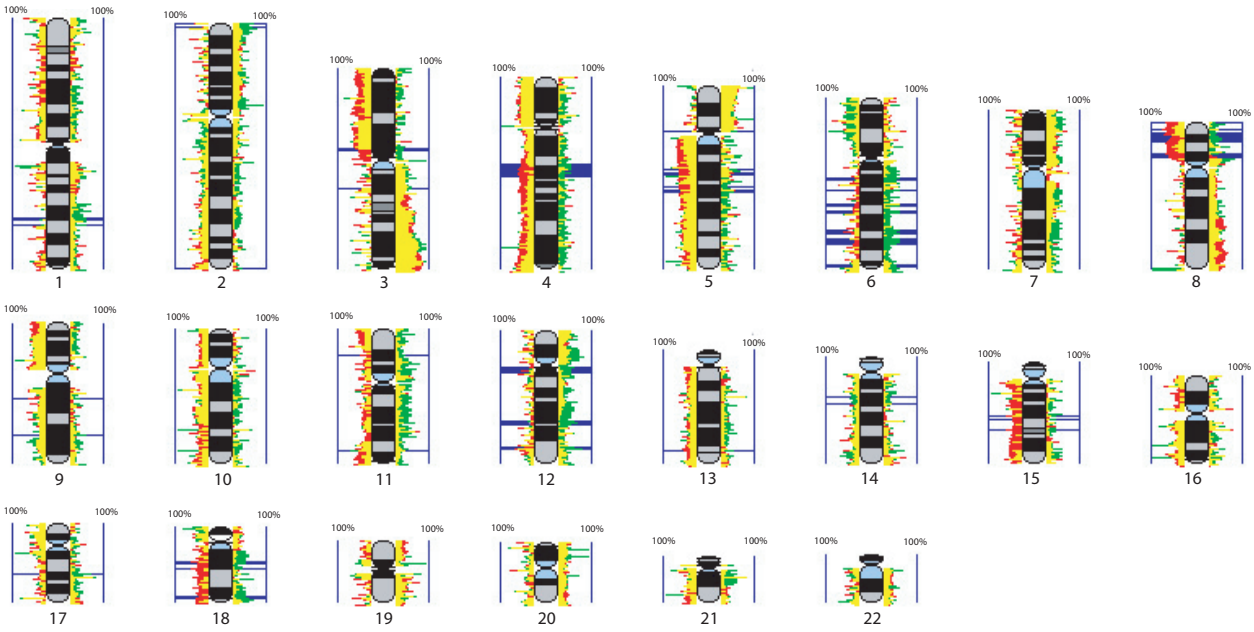
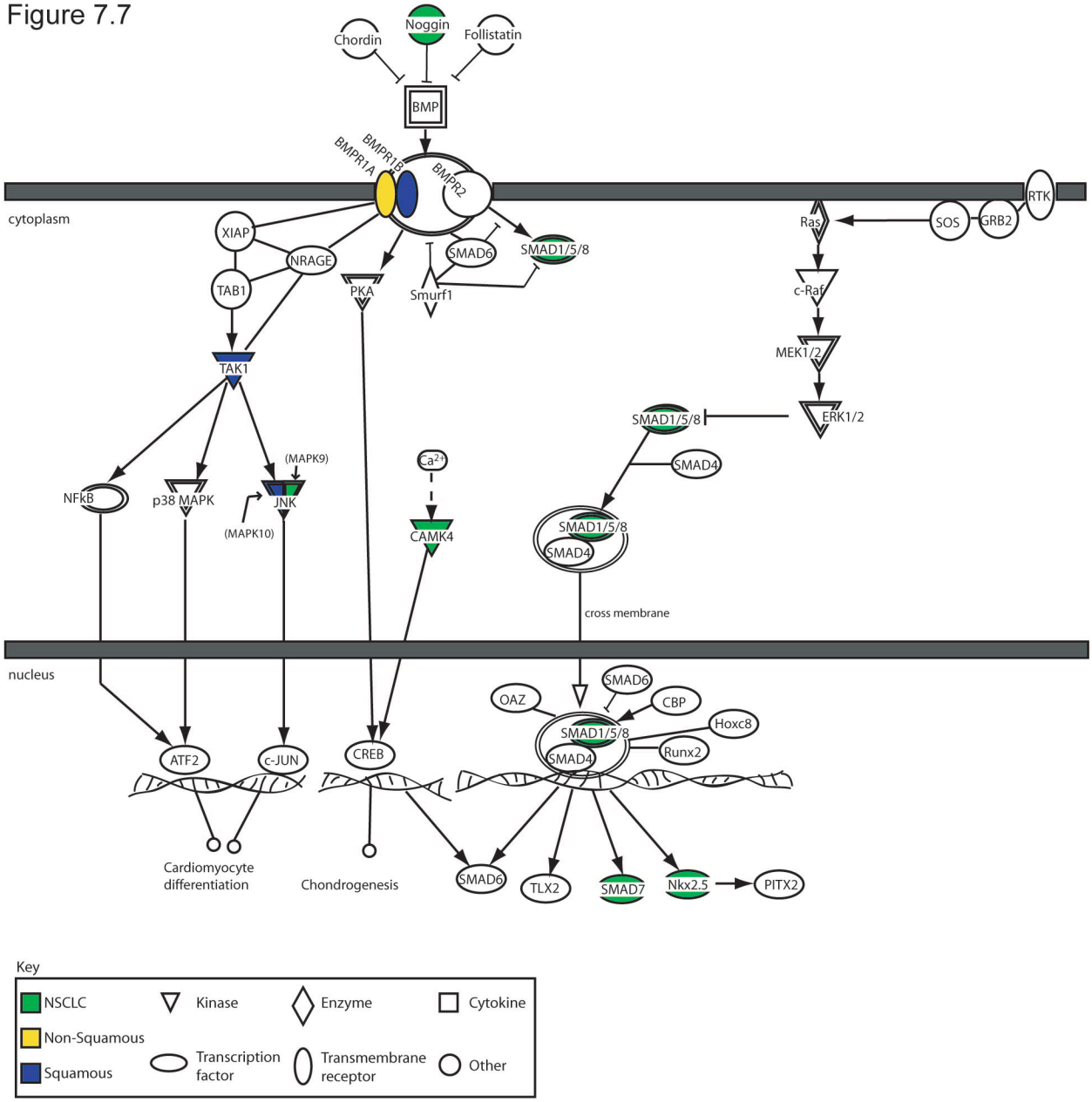

100%
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22

Figure 7.7 – Disruption of BMP pathway signaling in clinical NSCLC. All coloured genes exhibited differential activation in cases recurring following treatment with surgery and chemotherapy, with the majority affected by genomic loss in non-recurring specimens. Yellow pathway members were uncovered by analysis of non-squamous tumours, blue by analysis of squamous tumours, and green by combined analysis of NSCLC. Where arrows exist on lines between genes, an activating relationship is indicated (with the arrow depicting the direction of activation). Where bars exist at the end of a line between genes, an inhibitory relationship is inferred. An unannotated line between genes indicates a known interaction. All double-bordered pathway members were comprised of multiple genes (e.g. JNK, which included genes activated in both NSCLC and squamous specimens).

Figure 7.7



Key

NSCLC
Non-Squamous
Squamous
Kinase
Transcription factor
Enzyme
Transmembrane receptor
Cytokine
Other

Table 7.3 – Genes significantly associated with recurrence in NSCLC by genomic and gene expression.

Gene	Chromosome band	Relative Activation*	MannWhitney *U* test
CTSO	4q32.1	Recurring tumours	0.016
SHOC2	10q25.2	Recurring tumours	0.016
NRAP	10q25.3	Recurring tumours	0.016
HABP2	10q25.3	Recurring tumours	0.032
PNLIPRP3	10q25.3	Recurring tumours	0.032
C10orf90	10q26.2	Recurring tumours	0.032
SPRED1	15q14	Recurring tumours	0.016
KCTD19	16q22.1	Recurring tumours	0.048

*No gene candidates exhibited relative activation in non-recurring cases.

Table 7.4 – Features of NSCLC cell lines (n = 49).

Cell Line Name	Histological Subtype	Log(IC_{50}) Value*	
		***Cis*-platinum**	**Vinorelbine**
A549	Adenocarcinoma	0.455	*1.603*
Calu-1	Squamous (Epidermoid Carcinoma)	0.825	*0.212*
Calu-3	Adenocarcinoma	0.279	*1.686*
Calu-6	Squamous (Anaplastic Carcinoma)	*-0.043*	*0.340*
H1155	Large Cell Carcinoma	0.176	*0.505*
H1299	Large Cell Carcinoma	*0.017*	*0.892*
H1355	Adenocarcinoma	0.670	*1.149*
H1437	Adenocarcinoma	**1.204**	*1.114*
H157	Squamous Cell Carcinoma	0.371	*-0.420*
H1648	Adenocarcinoma	0.748	*0.140*
H1650	Adenocarcinoma (Bronchioloalveolar Carcinoma)	0.538	*0.477*
H1666	Adenocarcinoma (Bronchioloalveolar Carcinoma)	0.459	*0.780*
H1770	Neuroendocrine	0.279	*-0.178*
H1819	Adenocarcinoma	**0.968**	*-0.854*
H1975	Adenocarcinoma	0.565	*0.442*
H1993	Adenocarcinoma	0.802	*0.602*
H2009	Adenocarcinoma	0.740	*0.886*
H2085	Adenocarcinoma	**0.991**	*1.569*
H2087	Adenocarcinoma	*-0.208*	*0.371*
H2122	Adenocarcinoma	0.613	*0.204*
H2126	Adenocarcinoma	0.602	*1.079*
H2347	Adenocarcinoma	0.544	*0.991*
H2882	Non-small Cell Lung Cancer	0.255	*0.643*
H322	Adenocarcinoma (Bronchioloalveolar Carcinoma)	**1.267**	*0.462*
H3255	Adenocarcinoma	0.303	*0.267*
H358	Adenocarcinoma (Bronchioloalveolar Carcinoma)	0.611	*0.921*
H441	Adenocarcinoma	0.562	*0.778*
H460	Large Cell Carcinoma	*-0.310*	*1.041*
H820	Large Cell Carcinoma	0.127	*0.484*
HCC1195	Adenocarcinoma	0.771	*1.220*
HCC1359	Large Cell Carcinoma	0.863	*0.756*
HCC15	Squamous Cell Carcinoma	0.785	*0.658*
HCC193	Adenocarcinoma	**1.079**	*1.114*
HCC2279	Adenocarcinoma	**0.932**	*1.906*
HCC2935	Adenocarcinoma	**1.638**	*1.863*
HCC366	Adenocarcinoma	0.602	*-0.796*

Cell Line Name	Histological Subtype	Log(IC_{50}) Value*	
		***Cis*-platinum**	**Vinorelbine**
HCC4006	Adenocarcinoma	**1.143**	*0.685*
HCC44	Non-small Cell Lung Cancer	0.643	*1.000*
HCC461	Adenocarcinoma	**0.958**	*1.556*
HCC78	Adenocarcinoma	*-0.187*	*0.581*
HCC827	Adenocarcinoma	0.431	*0.556*
HCC95	Squamous Cell Carcinoma	0.342	*0.869*
H2887	Non-small Cell Lung Cancer	**1.041**	**2.633**
H1395	Adenocarcinoma	0.755	**3.000**
H1781	Adenocarcinoma (Bronchioloalveolar Carcinoma)	*-0.721*	**3.000**
H2073	Adenocarcinoma	0.204	**3.000**
HCC1171	Non-small Cell Lung Cancer	**0.944**	**3.000**
HCC4011	Non-small Cell Lung Cancer	0.775	**3.000**
HCC515	Adenocarcinoma	0.334	**3.000**

*Bold values indicate resistance to drug, italicized values indicate sensitivity.

Table 7.5. Genes associated with recurrence following surgery and adjuvant chemotherapy in non-squamous tumours and with CDDP-resistance in lung cancer cell lines.

Gene Symbol	Chromosome Band	Relative Activation
LOC402644	7p15.1	Recurring tumours
CREB5	7p15.1	Recurring tumours
CPVL	7p15.1	Recurring tumours
CHN2	7p15.1	Recurring tumours
WIPF3	7p15.1	Recurring tumours
SCRN1	7p15.1	Recurring tumours
FKBP14	7p15.1	Recurring tumours
PLEKHA8	7p15.1	Recurring tumours
ZNRF2	7p15.1	Recurring tumours
NOD1	7p15.1	Recurring tumours
GGCT	7p15.1	Recurring tumours
GARS	7p15.1	Recurring tumours
CRHR2	7p15.1	Recurring tumours
INMT	7p15.1	Recurring tumours
FLJ22374	7p15.1	Recurring tumours
AQP1	7p15.1	Recurring tumours
GHRHR	7p15.1	Recurring tumours
NEUROD6	7p15.1	Recurring tumours
CCDC129	7p15.1	Recurring tumours
C7orf16	7p15.1	Recurring tumours
PDE1C	7p14.3	Recurring tumours
LSM5	7p14.3	Recurring tumours
KIAA0241	7p14.3	Recurring tumours
KBTBD2	7p14.3	Recurring tumours
FKBP9	7p14.3	Recurring tumours
NT5C3	7p14.3	Recurring tumours
RP9	7p14.3	Recurring tumours
BBS9	7p14.3	Recurring tumours
BMPER	7p14.3	Recurring tumours
NPSR1	7p14.3	Recurring tumours
DPY19L1	7p14.3	Recurring tumours
HERPUD2	7p14.2	Recurring tumours

Gene Symbol	Chromosome Band	Relative Activation
KIAA0895	7p14.2	Recurring tumours
ANLN	7p14.2	Recurring tumours
ELMO1	7p14.2	Recurring tumours
GPR141	7p14.1	Recurring tumours
SFRP4	7p14.1	Recurring tumours
EPDR1	7p14.1	Recurring tumours
STARD3NL	7p14.1	Recurring tumours
TARP	7p14.1	Recurring tumours
AMPH	7p14.1	Recurring tumours
FAM183B	7p14.1	Recurring tumours
POU6F2	7p14.1	Recurring tumours
C7orf36	7p14.1	Recurring tumours
C7orf10	7p14.1	Recurring tumours
INHBA	7p14.1	Recurring tumours
GLI3	7p14.1	Recurring tumours
TNS3	7p12.3	Recurring tumours
LOC401335	7p12.3	Recurring tumours
PKD1L1	7p12.3	Recurring tumours
FLJ21075	7p12.3	Recurring tumours
HUS1	7p12.3	Recurring tumours
SUNC1	7p12.3	Recurring tumours
C7orf57	7p12.3	Recurring tumours
UPP1	7p12.3	Recurring tumours
ABCA13	7p12.3	Recurring tumours
VWC2	7p12.3-p12.2	Recurring tumours
FIGNL1	7p12.2	Recurring tumours
VSTM2A	7p11.2	Recurring tumours
EGFR	7p11.2	Recurring tumours

Figure 7.8

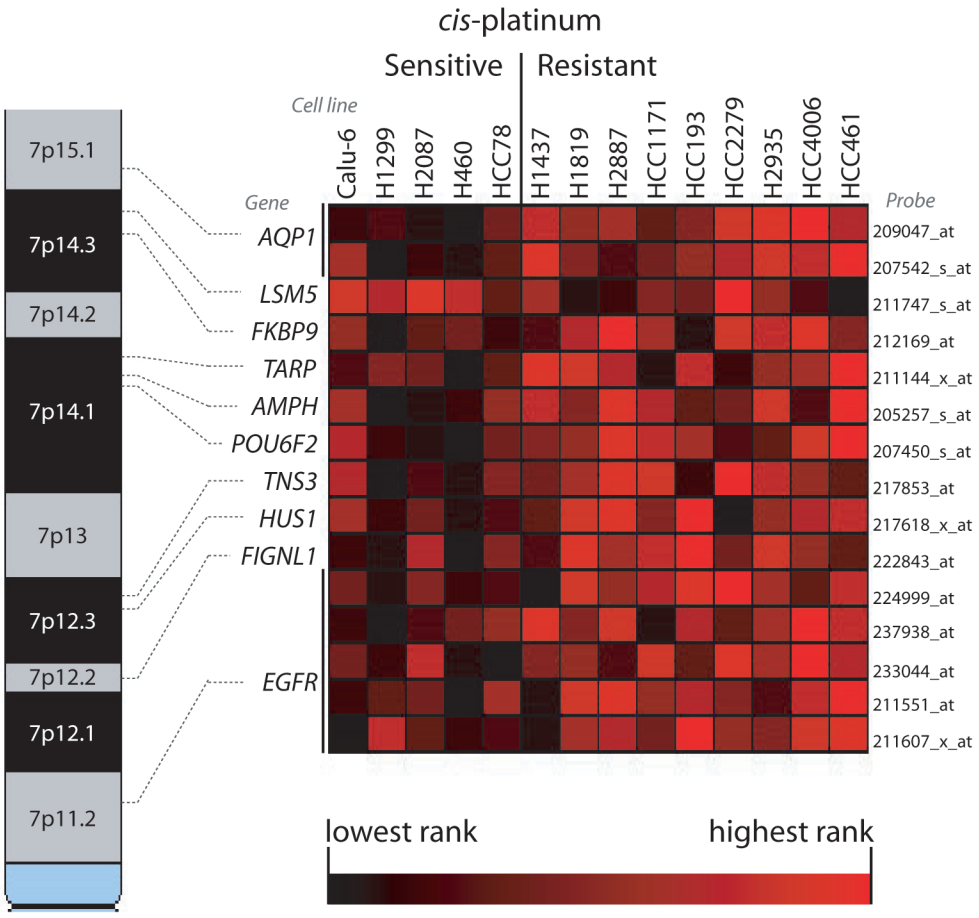


Figure 7.8 – Gene expression data for response-associated genes in non-squamous lung cancer cell lines. Matched gene expression profiles were available for 14/16 lines that were classified as sensitive or resistant to *cis*-platinum (five and nine lines respectively). Genomic alterations associated with chemoresponse in non-squamous clinical samples and lung cell lines were predominantly located on chromosome 7. Ten genes from this region were also found to have significantly different expression levels based on chemoresponse in cells. The chromosome band where each gene is located is indicated on the left. Gene expression microarray probe names are listed on the right, gene names on the left, and individual cell lines on the top of the heatmap display. Colourimetric values for each probe were assigned based on rank across the 14 lines studies, with the lowest intensity assigned to black and the highest to red.

7.6. References

Audebert, M., B. Salles & P. Calsou, 2004. Involvement of poly(ADP-ribose) polymerase-1 and XRCC1/DNA ligase III in an alternative route for DNA double-strand breaks rejoining. *J Biol Chem,* 279(53), 55117-26.

Brandt, P. D., C. E. Helt, P. C. Keng & R. A. Bambara, 2006. The Rad9 protein enhances survival and promotes DNA repair following exposure to ionizing radiation. *Biochem Biophys Res Commun,* 347(1), 232-7.

Burma, S., B. P. Chen & D. J. Chen, 2006. Role of non-homologous end joining (NHEJ) in maintaining genomic integrity. *DNA Repair (Amst),* 5(9-10), 1042-8.

Campbell, J. M., W. W. Lockwood, T. P. Buys, R. Chari, B. P. Coe, S. Lam & W. L. Lam, 2008. Integrative genomic and gene expression analysis of chromosome 7 identified novel oncogene loci in non-small cell lung cancer. *Genome,* 51(12), 1032-9.

Cappuzzo, F., C. Ligorio, L. Toschi, E. Rossi, R. Trisolini, D. Paioli, E. Magrini, G. Finocchiaro, S. Bartolini, A. Cancellieri, F. R. Hirsch, L. Crino & M. Varella-Garcia, 2007. EGFR and HER2 gene copy number and response to first-line chemotherapy in patients with advanced non-small cell lung cancer (NSCLC). *J Thorac Oncol,* 2(5), 423-9.

Ceppi, P., M. Volante, S. Novello, I. Rapa, K. D. Danenberg, P. V. Danenberg, A. Cambieri, G. Selvaggi, S. Saviozzi, R. Calogero, M. Papotti & G. V. Scagliotti, 2006. ERCC1 and RRM1 gene expressions but not EGFR are predictive of shorter survival in advanced non-small-cell lung cancer treated with cisplatin and gemcitabine. *Ann Oncol,* 17(12), 1818-25.

Chi, B., R. J. DeLeeuw, B. P. Coe, R. T. Ng, C. MacAulay & W. L. Lam, 2008. MD-SeeGH: a platform for integrative analysis of multi-dimensional genomic data. *BMC Bioinformatics,* In press.

Chin, T. M., M. P. Quinlan, A. Singh, L. V. Sequist, T. J. Lynch, D. A. Haber, S. V. Sharma & J. Settleman, 2008. Reduced Erlotinib sensitivity of epidermal growth factor receptor-mutant non-small cell lung cancer following cisplatin exposure: a cell culture model of second-line erlotinib treatment. *Clin Cancer Res,* 14(21), 6867-76.

Chong, I. W., M. Y. Chang, H. C. Chang, Y. P. Yu, C. C. Sheu, J. R. Tsai, J. Y. Hung, S. H. Chou, M. S. Tsai, J. J. Hwang & S. R. Lin, 2006. Great potential of a panel of multiple hMTH1, SPD, ITGA11 and COL11A1 markers for diagnosis of patients with non-small cell lung cancer. *Oncol Rep,* 16(5), 981-8.

Di Renzo, F., L. Doneda, E. Menegola, M. Sardella, G. De Vecchi, P. Collini, F. Spreafico, F. Fossati-Bellani, E. Giavini, P. Radice & D. Perotti, 2006. The murine Pou6f2 gene is temporally and spatially regulated during kidney embryogenesis and its human homolog is overexpressed in a subset of Wilms tumors. *J Pediatr Hematol Oncol,* 28(12), 791-7.

Duesberg, P., 2005. Does aneuploidy or mutation start cancer? *Science,* 307(5706), 41.

Duesberg, P., R. Li, R. Sachs, A. Fabarius, M. B. Upender & R. Hehlmann, 2007. Cancer drug resistance: the central role of the karyotype. *Drug Resist Updat,* 10(1-2), 51-8.

Eckstein, N., K. Servan, L. Girard, D. Cai, G. von Jonquieres, U. Jaehde, M. U. Kassack, A. F. Gazdar, J. D. Minna & H. D. Royer, 2008. Epidermal growth factor receptor pathway analysis identifies amphiregulin as a key factor for cisplatin resistance of human breast cancer cells. *J Biol Chem,* 283(2), 739-50.

Epel, M., I. Carmi, S. Soueid-Baumgarten, S. K. Oh, T. Bera, I. Pastan, J. Berzofsky & Y. Reiter, 2008. Targeting TARP, a novel breast and prostate tumor-associated antigen, with T cell receptor-like human recombinant antibodies. *Eur J Immunol,* 38(6), 1706-20.

Faivre, S., G. Kroemer & E. Raymond, 2006. Current development of mTOR inhibitors as anticancer agents. *Nat Rev Drug Discov,* 5(8), 671-88.

Floyd, S., M. H. Butler, O. Cremona, C. David, Z. Freyberg, X. Zhang, M. Solimena, A. Tokunaga, H. Ishizu, K. Tsutsui & P. De Camilli, 1998. Expression of amphiphysin I, an autoantigen of paraneoplastic neurological syndromes, in breast cancer. *Mol Med,* 4(1), 29-39.

Garnis, C., W. W. Lockwood, E. Vucic, Y. Ge, L. Girard, J. D. Minna, A. F. Gazdar, S. Lam, C. MacAulay & W. L. Lam, 2006. High resolution analysis of non-small cell lung cancer cell lines by whole genome tiling path array CGH. *Int J Cancer,* 118(6), 1556-64.

Gazdar, A. F., 2007. DNA repair and survival in lung cancer--the two faces of Janus. *N Engl J Med,* 356(8), 771-3.

Gelse, K., C. Muhle, K. Knaup, B. Swoboda, M. Wiesener, F. Hennig, A. Olk & H. Schneider, 2008. Chondrogenic differentiation of growth factor-stimulated precursor cells in cartilage repair tissue is associated with increased HIF-1alpha activity. *Osteoarthritis Cartilage,* 16(12), 1457-65.

Hardwick, J. C., L. L. Kodach, G. J. Offerhaus & G. R. van den Brink, 2008. Bone morphogenetic protein signalling in colorectal cancer. *Nat Rev Cancer,* 8(10), 806-12.

Hoque, M. O., J. C. Soria, J. Woo, T. Lee, J. Lee, S. J. Jang, S. Upadhyay, B. Trink, C. Monitto, C. Desmaze, L. Mao, D. Sidransky & C. Moon, 2006. Aquaporin 1 is overexpressed in lung cancer and stimulates NIH-3T3 cell proliferation and anchorage-independent growth. *Am J Pathol,* 168(4), 1345-53.

Jemal, A., M. J. Thun, L. A. Ries, H. L. Howe, H. K. Weir, M. M. Center, E. Ward, X. C. Wu, C. Eheman, R. Anderson, U. A. Ajani, B. Kohler & B. K. Edwards, 2008. Annual Report to the Nation on the Status of Cancer, 1975-2005, Featuring Trends in Lung Cancer, Tobacco Use, and Tobacco Control. *J Natl Cancer Inst*.

Jiang, B. H. & L. Z. Liu, 2008. Role of mTOR in anticancer drug resistance: perspectives for improved drug treatment. *Drug Resist Updat,* 11(3), 63-76.

Jong, K., E. Marchiori, G. Meijer, A. V. Vaart & B. Ylstra, 2004. Breakpoint identification and smoothing of array comparative genomic hybridization data. *Bioinformatics,* 20(18), 3636-7.

Karamouzis, M. V., J. R. Grandis & A. Argiris, 2007. Therapies directed against epidermal growth factor receptor in aerodigestive carcinomas. *Jama,* 298(1), 70-82.

Karni, R., Y. Hippo, S. W. Lowe & A. R. Krainer, 2008. The splicing-factor oncoprotein SF2/ASF activates mTORC1. *Proc Natl Acad Sci U S A,* 105(40), 15323-7.

Katz, M., I. Amit, A. Citri, T. Shay, S. Carvalho, S. Lavi, F. Milanezi, L. Lyass, N. Amariglio, J. Jacob-Hirsch, N. Ben-Chetrit, G. Tarcic, M. Lindzen, R. Avraham, Y. C. Liao, P. Trusk, A. Lyass, G. Rechavi, N. L. Spector, S. H. Lo, F. Schmitt, S. S. Bacus & Y. Yarden, 2007. A reciprocal tensin-3-cten switch mediates EGF-driven mammary cell migration. *Nat Cell Biol,* 9(8), 961-9.

Khojasteh, M., W. L. Lam, R. K. Ward & C. MacAulay, 2005. A stepwise framework for the normalization of array CGH data. *BMC Bioinformatics,* 6, 274.

Kim, I. Y. & S. J. Kim, 2006. Role of bone morphogenetic proteins in transitional cell carcinoma cells. *Cancer Lett,* 241(1), 118-23.

Kinzel, B., J. Hall, F. Natt, J. Weiler & D. Cohen, 2002. Downregulation of Hus1 by antisense oligonucleotides enhances the sensitivity of human lung carcinoma cells to cisplatin. *Cancer,* 94(6), 1808-14.

Kitisin, K., T. Saha, T. Blake, N. Golestaneh, M. Deng, C. Kim, Y. Tang, K. Shetty, B. Mishra & L. Mishra, 2007. Tgf-Beta signaling in development. *Sci STKE,* 2007(399), cm1.

Kwei, K. A., Y. H. Kim, L. Girard, J. Kao, M. Pacyna-Gengelbach, K. Salari, J. Lee, Y. L. Choi, M. Sato, P. Wang, T. Hernandez-Boussard, A. F. Gazdar, I. Petersen, J. D. Minna & J. R. Pollack, 2008. Genomic profiling identifies TITF1 as a lineage-specific oncogene amplified in lung cancer. *Oncogene,* 27(25), 3635-40.

Langenfeld, E. M., Y. Kong & J. Langenfeld, 2006. Bone morphogenetic protein 2 stimulation of tumor growth involves the activation of Smad-1/5. *Oncogene,* 25(5), 685-92.

Michaelis, M., J. Bliss, S. C. Arnold, N. Hinsch, F. Rothweiler, H. E. Deubzer, O. Witt, K. Langer, H. W. Doerr, W. S. Wels & J. Cinatl, Jr., 2008. Cisplatin-resistant neuroblastoma cells express enhanced levels of epidermal growth factor receptor (EGFR) and are sensitive to treatment with EGFR-specific toxins. *Clin Cancer Res,* 14(20), 6531-7.

National Cancer Institute of Canada, 2007. *Canadian Cancer Statistics 2007,* Toronto, Canada.

Petersen, I., M. Bujard, S. Petersen, G. Wolf, A. Goeze, A. Schwendel, H. Langreck, K. Gellert, M. Reichel, K. Just, S. du Manoir, T. Cremer, M. Dietel & T. Ried, 1997. Patterns of chromosomal imbalances in adenocarcinoma and squamous cell carcinoma of the lung. *Cancer Res,* 57(12), 2331-5.

Ramaswamy, S., K. N. Ross, E. S. Lander & T. R. Golub, 2003. A molecular signature of metastasis in primary solid tumors. *Nat Genet,* 33(1), 49-54.

Ries, L. A., D. Melbert, M. Krapcho, D. G. Stinchcomb, N. Howlader, M. J. Horner, A. Mariotto, B. A. Miller, E. J. Feuer, S. F. Altekruse, D. R. Lewis, L. Clegg, M. P. Eisner, M. Reichman & B. K. Edwards (eds.), (2008). *SEER Cancer Statistics Review, 1975-2005,* National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/csr/1975_2005/, based on November 2007 SEER data submission, posted to the SEER web site, 2008.

Rosell, R., M. Cuello, F. Cecere, M. Santarpia, N. Reguart, E. Felip & M. Taron, 2006. Usefulness of predictive tests for cancer treatment. *Bull Cancer,* 93(8), E101-8.

Sabatini, D. M., 2006. mTOR and cancer: insights into a complex relationship. *Nat Rev Cancer,* 6(9), 729-34.

Sambrook, J., E. F. Fritsch & T. Maniatis, (1989). Molecular Cloning: A laboratory manual, Cold Spring Harbor: Cold Spring Harbor Laboratory.

Sok, J. C., F. M. Coppelli, S. M. Thomas, M. N. Lango, S. Xi, J. L. Hunt, M. L. Freilino, M. W. Graner, C. J. Wikstrand, D. D. Bigner, W. E. Gooding, F. B. Furnari & J. R. Grandis, 2006. Mutant epidermal growth factor receptor (EGFRvIII) contributes to head and neck cancer growth and resistance to EGFR targeting. *Clin Cancer Res,* 12(17), 5064-73.

Sun, X., F. Li, N. Sun, Q. Shukui, C. Baoan, F. Jifeng, C. Lu, L. Zuhong, C. Hongyan, C. Yuandong, J. Jiazhong & Z. Yingfeng, 2009. Polymorphisms in XRCC1 and XPG and response to platinum-based chemotherapy in advanced non-small cell lung cancer patients. *Lung Cancer*.

Travis, W. D., 2002. Pathology of lung cancer. *Clin Chest Med,* 23(1), 65-81, viii.

Vardhanabhuti, S., S. J. Blakemore, S. M. Clark, S. Ghosh, R. J. Stephens & D. Rajagopalan, 2006. A comparison of statistical tests for detecting differential expression using Affymetrix oligonucleotide microarrays. *Omics,* 10(4), 555-66.

Volkmer, E. & L. M. Karnitz, 1999. Human homologs of Schizosaccharomyces pombe rad1, hus1, and rad9 form a DNA damage-responsive protein complex. *J Biol Chem,* 274(2), 567-70.

Wang, K. K., N. Liu, N. Radulovich, D. A. Wigle, M. R. Johnston, F. A. Shepherd, M. D. Minden & M. S. Tsao, 2002. Novel candidate tumor marker genes for lung adenocarcinoma. *Oncogene,* 21(49), 7598-604.

Watson, S. K., R. J. deLeeuw, D. E. Horsman, J. A. Squire & W. L. Lam, 2007. Cytogenetically balanced translocations are associated with focal copy number alterations. *Hum Genet,* 120(6), 795-805.

Weir, B. A., M. S. Woo, G. Getz, S. Perner, L. Ding, R. Beroukhim, W. M. Lin, M. A. Province, A. Kraja, L. A. Johnson, K. Shah, M. Sato, R. K. Thomas, J. A. Barletta, I. B. Borecki, S. Broderick, A. C. Chang, D. Y. Chiang, L. R. Chirieac, J. Cho, Y. Fujii, A. F. Gazdar, T. Giordano, H. Greulich, M. Hanna, B. E. Johnson, M. G. Kris, A. Lash, L. Lin, N. Lindeman, E. R. Mardis, J. D. McPherson, J. D. Minna, M. B. Morgan, M. Nadel, M. B. Orringer, J. R. Osborne, B. Ozenberger, A. H. Ramos, J. Robinson, J. A. Roth, V. Rusch, H. Sasaki, F. Shepherd, C. Sougnez, M. R. Spitz, M. S. Tsao, D. Twomey, R. G. Verhaak, G. M. Weinstock, D. A. Wheeler, W. Winckler, A. Yoshizawa, S. Yu, M. F. Zakowski, Q. Zhang, D. G. Beer, Wistuba, II, M. A. Watson, L. A. Garraway, M. Ladanyi, W. D. Travis, W. Pao, M. A. Rubin, S. B. Gabriel, R. A. Gibbs, H. E. Varmus, R. K. Wilson, E. S. Lander & M. Meyerson, 2007. Characterizing the cancer genome in lung adenocarcinoma. *Nature,* 450(7171), 893-8.

8. SUMMARY[21]

8.1. Research summary

In Canada and the United States, non-small cell lung cancer (NSCLC) accounts for nearly as many deaths as do prostate, breast, and colorectal malignancies combined (Jemal et al. 2008; Canada 2007). This poor survival is largely attributable to the high metastatic potential of even small NSCLC tumours and the lack of sensitivity to existing systemic therapies. Over the last five years, improving technologies for analyzing DNA and gene expression changes in tumour genomes have been applied to better understand the molecular basis for disease phenotypes. Such analyses are meant to yield new tools for predicting tumour behaviour and to provide additional candidates for the next generation of targeted therapies, both of which could significantly improve outcomes.

In this thesis, we began by describing the development of comprehensive approaches to sample collection and genomic analysis for clinical lung cancer. We applied these approaches to identify novel oncogene candidates in lung tumours and in cancer cell lines. We then took advantage of high resolution genome profiling to comment on clonal relationships between multiple tumours from the same patient and to define genes driving chemoresistance during drug-selected clonal evolution. Finally, we applied integrative analysis of genomic and gene expression data from both clinical lesions and NSCLC cell lines to define genes and molecular signaling cascades contributing to tumour recurrence after treatment with surgery and adjuvant chemotherapy.

8.1.1. Optimizing approaches to sample handling and genomic profiling for clinical lung cancer

High resolution platforms for whole genome analysis represent a critical tool for discovering the causal gene changes underlying tumour behaviour. We have undertaken much work to optimize the manufacture and application of such a platform

[21] Text included in this chapter originally appeared in abstracts for the manuscripts used for chapters 2-7.

to clinical cancer specimens (Ishkanian et al. 2004; Watson et al. 2007; Watson et al. 2004; Buys et al. 2007b). The ability of the tiling-path BAC array to profile lower quality DNA samples has facilitated analysis of archived biopsy specimens, meaning that genomic data could be obtained for non-resected tissues such as early stage lesions and inoperable late stage tumours. In addition to optimizing microarray protocols and analytical approaches, we also optimized approaches to lung tumour collection to ensure that best quality DNA for a given case was available for analysis. Precise microdissection of tumour cells was also applied to minimize the impact of tissue heterogeneity on genomic signatures (Garnis et al. 2005a). This initial attention to both the development of a robust genome profiling technology and the collection of lung tumour DNA served as the foundation for the various molecular analyses described in subsequent chapters.

We next sought to demonstrate that the genomic approaches defined in Chapter 2 could effectively be applied to specific research questions. In chapter 3, we analyzed molecular alterations in premalignant lung lesions and in invasive disease (Garnis et al. 2005b). Despite the fact that pre-invasive lesions are challenging to isolate, and often yield insufficient DNA for the analysis of multiple loci, genomic profiling of such lesions is worthwhile since it can lead to the discovery of causal genetic alterations that might otherwise be masked by the gross instability associated with invasive tumours. We identified multiple early genetic events on chromosome 5p in lung cancer progression. Using a tiling-path chromosome 5p-specific genomic array, nine novel minimal regions of loss and gain were discovered in bronchial carcinoma *in situ* (CIS) specimens. Within these regions we identified two candidate genes novel to lung cancer. The 0.27 Mbp region at 5p15.2 contains a single gene, *Triple Functional Domain* (*TRIO*), which we determined to be differentially expressed in tumours. The 0.34 Mbp region at 5p13.2 contains *Glial Cell Line-Derived Neurotrophic Factor* (*GDNF*), which is a ligand for the *RET* oncogene product and is normally expressed during lung development (but absent in adult lung tissue). Our data showed not only that *GDNF* is overexpressed at the transcript level in squamous non-small-cell lung carcinoma, but also that the GDNF

protein is present in early-stage lesions. Reactivation of the fetal lung-expressed *GDNF* in early lesions and its amplification in CIS suggests an early role in tumourigenesis.

We also sought to confirm the utility of our genomic approaches in integrative analysis of lung cancer cell model systems and clinical lung tumours (Campbell et al. 2008). Several previous genome studies have identified recurring aberrations on chromosome 7 in NSCLC. The presence of recurring chromosome 7 alterations that do not span the well-studied oncogenes *EGFR* (at 7p11.2) and *MET* (at 7q31.2) has raised the possibility that additional genes on this chromosome contribute to tumourigenesis. In Chapter 4, we demonstrated that multiple loci on chromosome 7 are indeed amplified in NSCLC, and through integrative analysis of gene dosage alterations and parallel gene expression changes, we identified new lung cancer oncogene candidates, including *FTSJ2*, *NUDT1*, *TAF6*, and *POLR2J*. Activation of these key genes was confirmed in panels of clinical lung tumour tissue as compared to matched normal lung tissue.

8.1.2. Delineating clonal relationships and clonal evolution in lung cancer

Moving beyond gene discovery, we next applied fine-mapped genome alteration boundaries as signature markers for clonality. In Chapter 5, we used this approach to evaluate the clonal relationship between nodules for a rare case where a patient presented with a combined small cell lung carcinoma (SCLC), large cell neuroendocrine carcinoma (LCNEC) and adenocarcinoma (AC) (Buys et al. 2009). In two areas of SCLC distinguishable by divergent neuroendocrine markers expression (CD56 and chromogranin-A), the presence of identical genomic breakpoints and rearrangements indicated a common origin, with the presence of additional distinct genomic alterations in these two components indicating diverging clonal evolution. The absence of shared genome alteration features for the AC and LCNEC components suggested that these tumours evolved independently from the SCLC lesions.

In Chapter 6 we then applied our high resolution genome analysis to cell populations that were known to be clonally-related in order to characterize regions of genomic difference (Buys et al. 2007a). Previously, multidrug resistance (MDR) phenotypes have been attributed to the activity of ATP-binding cassette (ABC) transporters such as

P-glycoprotein (*ABCB1*). Earlier work has suggested that modulation of MDR may not necessarily be a single gene trait. To identify factors contributing to the emergence of MDR, we undertook integrative genomic analysis of the ovarian carcinoma cell line SKOV3 and a series of MDR derivatives of this line (SKVCRs). As resistance increased, comparative analysis of gene expression showed conspicuous activation of a network of genes in addition to *ABCB1*. Functional annotation and pathway analysis revealed that many of these genes were associated with the extracellular matrix and had previously been implicated in tumour invasion and cell proliferation. Further investigation by whole genome tiling-path array CGH suggested that changes in gene dosage were essential for the activation of several of these over-expressed genes. Remarkably, alignment of whole genome profiles for SKVCR lines revealed the emergence and decline of specific segmental DNA alterations. The most prominent alteration was a novel amplicon residing at 16p13 that encompassed the ABC transporter genes *ABCC1* and *ABCC6*. Loss of this amplicon in highly resistant SKVCR lines coincided with the emergence of a different amplicon at 7q21.12, which harbors *ABCB1*. Integrative analysis suggests then that multiple genes are activated during escalation of drug resistance, including a succession of ABC transporter genes and genes that may act synergistically with *ABCB1*.

8.2. Lung cancer pharmacogenomics

Finally, in Chapter 7, we examined a panel of ~200 resected early stage NSCLC tumours to identify genomic alterations associated with recurrence following treatment with an adjuvant *cis*-platinum/vinorelbine doublet. Post-treatment recurrence status for each case was determined based on a minimum follow-up of two years. Genomic alterations significantly associated with recurrence were initially defined in the subset of patients who were treated by both surgery and chemotherapy (n = 81). Significant alterations defined for patients treated by surgery alone were then subtracted, leaving those DNA changes more likely to mediate chemoresponse. This same analysis was repeated within the panel for squamous tumours and for non-squamous tumours. Independent analyses for NSCLC and separate histological subtypes indicated that retention/activation of BMP pathway signaling was significantly associated with post-

treatment recurrence. Additional analysis was undertaken in NSCLC cell lines which had known response to *cis*-platinum (CDDP) and vinorelbine. Combined genomic and gene expression analysis of tumours and cell lines revealed several associated several individual genes with post-treatment recurrence. We observed prominent activation of chromosome arm 7p in non-squamous tumours in the context of CDDP resistance, with several genes in this region apparently gained and over-expressed (including *EGFR*). Broad dysregulation of chromosome 7 as a means of activating several oncogenes was a phenomenon reported earlier in this thesis [Chapter 4] (Campbell et al. 2008) and our findings support that this is a critical alteration in lung cancer.

8.3. Conclusions

We have developed effective approaches to sample collection and genomic data analysis in clinical lung cancer. Specimen accrual, cancer cell microdissection, microarray production and usage, and integration of various molecular data have been optimized to uncover critical changes contributing to lung cancer phenotypes. Parallel analysis using genomic approaches defined in this thesis characterized recurring DNA copy number alterations in premalignant disease, revealing candidate biomarkers for disease progression. Use of these same approaches in a combined analysis of cell model systems and clinical lung tumours uncovered several oncogene candidates within a single chromosomal region (chromosome 7), suggesting that imbalance at this locus may contribute to tumourigenesis via several mechanisms.

Based on our subsequent analysis, we conclude that fine-mapped genomic changes have great utility as signature markers of clonality. We initially demonstrated that the presence of identical DNA alteration boundaries in multiple tumour specimens could be used to infer a shared clonal origin. This application of whole genome analysis shows how high resolution molecular tools may have great utility for resolving the origin of and delineating the clonal relationships between tumours from the same patient. We next applied this fine-mapping approach to define differences in cell populations already known to have a shared clonal origin, thus uncovering critical gene changes that could be associated with a specific cancer phenotype (in this case, chemoresistance). The

sum of this analysis was that the evolution of the MDR phenotype appears to be a dynamic, multi-genic process in the genomes of cancer cells.

We also used genomic data from lung tumours to draw multiple conclusions regarding the basis for disease recurrence following treatment with adjuvant chemotherapy. First, our discovery of recurrence-associated genome changes in pre-treatment specimens indicates that DNA changes mediating chemoresponse exist prominently in tumour cell populations before exposure to drug. These findings suggests that chemoresistance may not be mediated by a tiny subpopulation of resistant cells and that pro-resistance DNA changes may in fact be tied to earlier malignant processes. Second, the fact that some of the most striking recurrence-associated DNA alterations were identified when analysis was restricted within histological subgroups lends support to the idea that different cellular morphologies reflect differences in underlying tumour biology. The mounting evidence that these subgroups are driven by different molecular changes suggests that analyzing NSCLC as a single disease entity may be a flawed approach and that greater emphasis should placed on analysis within subgroups such as lung squamous cell carcinoma.

Finally, we gained several insights into the importance of initial experiment design and the need for ongoing audits of data and results. We learned that up-front knowledge of all the questions that will be applied to a given dataset is beneficial, since not all experimental results may be usable when additional patient demographic data are introduced (this is because some patients may be inevaluable based on these new criteria). This is important for tempering expectations for what can be learned from a given sample set. We also learned that ongoing evaluation of experimental data is essential. For example, the need to normalize for microarray hybridization artifacts arising due to poor quality DNA only became apparent after side-by-side examination of tumors from unrelated individuals revealed highly recurrent alterations. Without this discovery, we would not have taken steps to remove this bias and would have detected these alterations as genuine changes. Based on these sorts of challenges, we conclude that it may ultimately be more effective to analyze samples accrued from

existing clinical trials, where collection of demographic data and standardized approaches to patient follow-up are *de rigueur*.

8.4. Future directions

In terms of our analysis of chemoresponse in lung cancer, two series of experiments would logically follow from our results. The first would involve an expanded clinical analysis of our proposed biomarkers. For example, the predictive power of EGFR and BMP signaling pathway genomic dysregulation in the context of recurrence for non-squamous tumours will need to be measured against existing markers for recurrence following platinum-based therapy (e.g. ERCC1, RRM1, and BRCA1) (Rosell et al. 2006a; Rosell et al. 2006b). On a longer time-line and with expanded resources, we could evaluate our markers in a march larger patient cohort to improve the statistical power of our analyses. In this scenario, we would restrict enrollment to only one lung cancer subtype (e.g. lung adenocarcinoma). This step would be taken to maximize our evaluable data set, given our finding that gene alterations associated with drug response are different between tumour histological subgroups. (Selection for this group would also likely be refined to include only current and former smokers, since there is increasing evidence that tumours from non-smokers are driven by distinct molecular changes [Sun et al. 2007]). DNA and RNA markers for recurrence that were derived by analysis of several hundred lung cancer cases would have a strong rationale use in a clinical trial.

The second set of experiments logically following our chemoresponse findings would involve further characterization of the mechanism by which candidate genes govern chemoresistance. This work would be most effective where candidates had been confirmed in the larger sample set described above. Initially, we would manipulate expression of resistance genes in lung cancer cells with known sensitivities to CDDP and vinorelbine. Significant changes in drug IC_{50} values based on this perturbation would confirm a direct role for candidates in response. Then, once a gene had been strongly linked to resistance against a specific drug, we would assess the downstream effects of its activation. The assays used for these experiments would be chosen based

on the known action of the candidate gene. For example, if the candidate was a transcription factor, we would identify its downstream targets by manipulating its expression . The logical step after this would be to characterize how these downstream targets might contribute to the resistance phenotype. Taken together, these experiments would define the processes underlying lung cancer chemoresistance and provide druggable targets for new forms of chemotherapy.

Another project that could logically follow this thesis in the short term would involve analysis of a panel of lung cancer cases from patients that presented with multiple tumours (this panel is actually being assembled right now). These cases could be used to test our ability to define clonal relationships using fine-mapped genomic alteration boundaries. For example, we could check to see whether cases we defined as clonally-related were strongly associated with the development of distant metastases (as would be expected if we had truly described pulmonary mets). Complex algorithms for defining clonality based on high resolution genomic data are being developed and we are already working to adapt these to our analysis of such cases (Begg et al. 2007).

Finally, a highly practical application based on the work in this thesis would be the sharing of our rigorous standards for clinical collection, handling, and storage of lung tumour specimens with other institutes. The fruits of our experience in this area could be adopted across the country to ensure that high quality tissues were available for molecular studies on a pan-Canadian scale.

8.5. Significance of work

We have shown that it is feasible to apply whole genome profiling technologies to clinical lung cancer specimens, demonstrating particularly that meaningful genomic data may be obtained from a large panel of archived tumour tissues. We have also described a genomic approach for defining clonal relationships between tumours, providing a tool that could impact on patient management strategies. Finally, we report that chemoresponse in different histological subtypes of lung cancer may be driven by different gene alterations, a finding that could have a significant impact on the way that NSCLC is managed. The sum of this work has been the discovery of critical gene

changes associated with cancer phenotypes such as tumour invasion and chemoresistance. Upon further validation, these candidate genes will have utility as prognostic biomarkers and may also be useful as targets for novel therapeutics that will improve lung cancer survival.

8.6. References

Begg, C. B., K. Eng, A. Olshen & W. S. Venkatraman, 2007. Statistical Evaluation of Evidence for Clonal Allelic Alterations in array-CGH Experiments. *Memorial Sloan-Kettering Cancer Center Department of Epidemiology and Biostatistics - Working Paper Series,* 13, 1-31.

Buys, T. P., R. Chari, E. H. Lee, M. Zhang, C. MacAulay, S. Lam, W. L. Lam & V. Ling, 2007a. Genetic changes in the evolution of multidrug resistance for cultured human ovarian cancer cells. *Genes Chromosomes Cancer,* 46(12), 1069-79.

Buys, T. P., I. M. Wilson, B. P. Coe, W. W. Lockwood, J. J. Davies, R. Chari, R. DeLeeuw, A. Shadeo, C. MacAulay & W. L. Lam, (2007b). Key Features of BAC Array Production and Usage, in *DNA Microarrays (Methods Express Series)*, ed. M. Schena Bloxham: Scion Publishing, Ltd.

Buys, T. P. H., S. Aviel-Ronen, T. K. Waddell, W. L. Lam & M. S. Tsao, 2009. Defining Genomic Alteration Boundaries for a Combined Small Cell and Non-Small Cell Lung Carcinoma. *J Thorac Oncol,* in press.

Campbell, J. M., W. W. Lockwood, T. P. Buys, R. Chari, B. P. Coe, S. Lam & W. L. Lam, 2008. Integrative genomic and gene expression analysis of chromosome 7 identified novel oncogene loci in non-small cell lung cancer. *Genome,* 51(12), 1032-9.

Garnis, C., B. P. Coe, S. L. Lam, C. Macaulay & W. L. Lam, 2005a. High-resolution array CGH increases heterogeneity tolerance in the analysis of clinical samples. *Genomics,* 85(6), 790-3.

Garnis, C., J. J. Davies, T. P. Buys, M. S. Tsao, C. Macaulay, S. Lam & W. L. Lam, 2005b. Chromosome 5p aberrations are early events in lung cancer: implication of glial cell line-derived neurotrophic factor in disease progression. *Oncogene,* 24(30).

Ishkanian, A. S., C. A. Malloff, S. K. Watson, R. J. DeLeeuw, B. Chi, B. P. Coe, A. Snijders, D. G. Albertson, D. Pinkel, M. A. Marra, V. Ling, C. MacAulay & W. L. Lam, 2004. A tiling resolution DNA microarray with complete coverage of the human genome. *Nat Genet,* 36(3), 299-303.

Jemal, A., M. J. Thun, L. A. Ries, H. L. Howe, H. K. Weir, M. M. Center, E. Ward, X. C. Wu, C. Eheman, R. Anderson, U. A. Ajani, B. Kohler & B. K. Edwards, 2008. Annual Report to the Nation on the Status of Cancer, 1975-2005, Featuring Trends in Lung Cancer, Tobacco Use, and Tobacco Control. *J Natl Cancer Inst*.

National Cancer Institute of Canada, 2007. *Canadian Cancer Statistics 2007,* Toronto, Canada.

Rosell, R., F. Cecere, M. Santarpia, N. Reguart & M. Taron, 2006a. Predicting the outcome of chemotherapy for lung cancer. *Curr Opin Pharmacol,* 6(4), 323-31.

Rosell, R., M. Cuello, F. Cecere, M. Santarpia, N. Reguart, E. Felip & M. Taron, 2006b. Usefulness of predictive tests for cancer treatment. *Bull Cancer,* 93(8), E101-8.

Sun, S., J. H. Schiller & A. F. Gazdar, 2007. Lung cancer in never smokers--a different disease. *Nat Rev Cancer,* 7(10), 778-90.

Watson, S. K., R. J. deLeeuw, D. E. Horsman, J. A. Squire & W. L. Lam, 2007. Cytogenetically balanced translocations are associated with focal copy number alterations. *Hum Genet,* 120(6), 795-805.

Watson, S. K., R. J. DeLeeuw, A. S. Ishkanian, C. A. Malloff & W. L. Lam, 2004. Methods for high throughput validation of amplified fragment pools of BAC DNA for constructing high resolution CGH arrays. *BMC Genomics,* 5(1), 6.

Appendix 1 – SUPPLEMENTARY DATA FOR CHAPTER 7

Supplementary Table 7.S1 – Genes significantly associated with recurrence following treatment with surgery and adjuvant chemotherapy based on combined analysis of NSCLC tumours (n = 81).

Gene Name	Chromosome	Group with relative activation
C1orf111	1	Non-recurring tumours
C1orf226	1	Non-recurring tumours
LOC100133912	1	Non-recurring tumours
NOS1AP	1	Non-recurring tumours
SH2D1B	1	Non-recurring tumours
ACCN5	4	Recurring tumours
CTSO	4	Recurring tumours
GUCY1B3	4	Recurring tumours
HHIP	4	Recurring tumours
IL15	4	Recurring tumours
INPP4B	4	Recurring tumours
SMAD1	4	Recurring tumours
TBC1D9	4	Recurring tumours
TDO2	4	Recurring tumours
ZNF330	4	Recurring tumours
ABLIM3	5	Recurring tumours
ANKRA2	5	Recurring tumours
AP3S1	5	Recurring tumours
APC	5	Recurring tumours
ATG12	5	Recurring tumours
BNIP1	5	Recurring tumours
C5orf13	5	Recurring tumours
C5orf41	5	Recurring tumours
CAMK4	5	Recurring tumours
CANX	5	Recurring tumours
CCDC112	5	Recurring tumours
CCDC99	5	Recurring tumours
CDO1	5	Recurring tumours
CEP120	5	Recurring tumours
CLK4	5	Recurring tumours
CNOT6	5	Recurring tumours
COL23A1	5	Recurring tumours
COMMD10	5	Recurring tumours
CPEB4	5	Recurring tumours
CPLX2	5	Recurring tumours
CSNK1G3	5	Recurring tumours
DCP2	5	Recurring tumours
DOCK2	5	Recurring tumours
DRD1	5	Recurring tumours

Gene Name	Chromosome	Group with relative activation
DTWD2	5	Recurring tumours
DUSP1	5	Recurring tumours
EFNA5	5	Recurring tumours
EPB41L4A	5	Recurring tumours
FAM170A	5	Recurring tumours
FAM44B	5	Recurring tumours
FBXL17	5	Recurring tumours
FBXW11	5	Recurring tumours
FEM1C	5	Recurring tumours
FER	5	Recurring tumours
FGF18	5	Recurring tumours
FOXI1	5	Recurring tumours
GABRP	5	Recurring tumours
GFPT2	5	Recurring tumours
HAPLN1	5	Recurring tumours
HMHB1	5	Recurring tumours
HMP19	5	Recurring tumours
HNRNPH1	5	Recurring tumours
HSD17B4	5	Recurring tumours
KCNIP1	5	Recurring tumours
KCNMB1	5	Recurring tumours
KCNN2	5	Recurring tumours
LCP2	5	Recurring tumours
LOC100131897	5	Recurring tumours
LOC133874	5	Recurring tumours
LOX	5	Recurring tumours
LVRN	5	Recurring tumours
MAN2A1	5	Recurring tumours
MAPK9	5	Recurring tumours
MCC	5	Recurring tumours
MGAT1	5	Recurring tumours
MSX2	5	Recurring tumours
NKX2-5	5	Recurring tumours
NPM1	5	Recurring tumours
NR3C1	5	Recurring tumours
ODZ2	5	Recurring tumours
OR2Y1	5	Recurring tumours
PAM	5	Recurring tumours
PANK3	5	Recurring tumours
PGGT1B	5	Recurring tumours
PJA2	5	Recurring tumours
PRR16	5	Recurring tumours
RANBP17	5	Recurring tumours
RARS	5	Recurring tumours
RASGEF1C	5	Recurring tumours
REEP5	5	Recurring tumours
RGNEF	5	Recurring tumours
RNF130	5	Recurring tumours

Gene Name	Chromosome	Group with relative activation
RUFY1	5	Recurring tumours
SCGB3A1	5	Recurring tumours
SEMA6A	5	Recurring tumours
SFXN1	5	Recurring tumours
SH3PXD2B	5	Recurring tumours
SH3TC2	5	Recurring tumours
SLC25A46	5	Recurring tumours
SLIT3	5	Recurring tumours
SNCAIP	5	Recurring tumours
SRFBP1	5	Recurring tumours
SRP19	5	Recurring tumours
STC2	5	Recurring tumours
STK10	5	Recurring tumours
THOC3	5	Recurring tumours
TICAM2	5	Recurring tumours
TMED7	5	Recurring tumours
TRIM36	5	Recurring tumours
TSLP	5	Recurring tumours
TSSK1B	5	Recurring tumours
UBTD2	5	Recurring tumours
UTP15	5	Recurring tumours
VCAN	5	Recurring tumours
WDR36	5	Recurring tumours
XRCC4	5	Recurring tumours
YTHDC2	5	Recurring tumours
ZFP62	5	Recurring tumours
ZNF474	5	Recurring tumours
KIAA0196	8	Non-recurring tumours
MAL2	8	Non-recurring tumours
MTSS1	8	Non-recurring tumours
NSMCE2	8	Non-recurring tumours
SQLE	8	Non-recurring tumours
TRIB1	8	Non-recurring tumours
ZHX2	8	Non-recurring tumours
ZNF572	8	Non-recurring tumours
CSMD1	8	Recurring tumours
ABLIM1	10	Recurring tumours
ADRA2A	10	Recurring tumours
ADRB1	10	Recurring tumours
AFAP1L2	10	Recurring tumours
ATRNL1	10	Recurring tumours
C10orf118	10	Recurring tumours
C10orf81	10	Recurring tumours
C10orf82	10	Recurring tumours
C10orf90	10	Recurring tumours
C10orf96	10	Recurring tumours
CASP7	10	Recurring tumours
DKK1	10	Recurring tumours

Gene Name	Chromosome	Group with relative activation
DUSP5	10	Recurring tumours
FAM160B1	10	Recurring tumours
GFRA1	10	Recurring tumours
HABP2	10	Recurring tumours
HECTD2	10	Recurring tumours
HPSE2	10	Recurring tumours
HSPA12A	10	Recurring tumours
MXI1	10	Recurring tumours
NRAP	10	Recurring tumours
PDCD4	10	Recurring tumours
PNLIP	10	Recurring tumours
PNLIPRP1	10	Recurring tumours
PNLIPRP2	10	Recurring tumours
PNLIPRP3	10	Recurring tumours
PPP1R3C	10	Recurring tumours
PRKG1	10	Recurring tumours
RBM20	10	Recurring tumours
SHOC2	10	Recurring tumours
SMC3	10	Recurring tumours
SMNDC1	10	Recurring tumours
SORCS1	10	Recurring tumours
SORCS3	10	Recurring tumours
TCF7L2	10	Recurring tumours
TDRD1	10	Recurring tumours
TNKS2	10	Recurring tumours
TRUB1	10	Recurring tumours
VTI1A	10	Recurring tumours
VWA2	10	Recurring tumours
OR10A5	11	Recurring tumours
OR2AG1	11	Recurring tumours
OR2AG2	11	Recurring tumours
OR6A2	11	Recurring tumours
FBXL14	12	Non-recurring tumours
WNT5B	12	Non-recurring tumours
SPRED1	15	Recurring tumours
A2BP1	16	Recurring tumours
ADCY7	16	Recurring tumours
ATP6V0D1	16	Recurring tumours
BRD7	16	Recurring tumours
C16orf61	16	Recurring tumours
C16orf70	16	Recurring tumours
CBFB	16	Recurring tumours
CDH16	16	Recurring tumours
CENPN	16	Recurring tumours
CES2	16	Recurring tumours
CES3	16	Recurring tumours
CLEC3A	16	Recurring tumours
E2F4	16	Recurring tumours

Gene Name	Chromosome	Group with relative activation
ELMO3	16	Recurring tumours
EXOC3L	16	Recurring tumours
FAM96B	16	Recurring tumours
FHOD1	16	Recurring tumours
FLJ37464	16	Recurring tumours
HSD11B2	16	Recurring tumours
KCTD19	16	Recurring tumours
LOC653319	16	Recurring tumours
LRRC29	16	Recurring tumours
LRRC36	16	Recurring tumours
NKD1	16	Recurring tumours
PDP2	16	Recurring tumours
PLEKHG4	16	Recurring tumours
RRAD	16	Recurring tumours
SLC9A5	16	Recurring tumours
TMEM208	16	Recurring tumours
TPPP3	16	Recurring tumours
WWOX	16	Recurring tumours
ZDHHC1	16	Recurring tumours
ANKFN1	17	Recurring tumours
C17orf67	17	Recurring tumours
CUEDC1	17	Recurring tumours
DGKE	17	Recurring tumours
DYNLL2	17	Recurring tumours
EPX	17	Recurring tumours
KIAA1303	17	Recurring tumours
LPO	17	Recurring tumours
MKS1	17	Recurring tumours
MPO	17	Recurring tumours
MRPS23	17	Recurring tumours
MSI2	17	Recurring tumours
NOG	17	Recurring tumours
OR4D1	17	Recurring tumours
OR4D2	17	Recurring tumours
SFRS1	17	Recurring tumours
TRIM25	17	Recurring tumours
VEZF1	17	Recurring tumours
ACAA2	18	Recurring tumours
ATP9B	18	Recurring tumours
CTDP1	18	Recurring tumours
GALR1	18	Recurring tumours
HDHD2	18	Recurring tumours
IER3IP1	18	Recurring tumours
KATNAL2	18	Recurring tumours
MAPK4	18	Recurring tumours
MBP	18	Recurring tumours
MYO5B	18	Recurring tumours
NFATC1	18	Recurring tumours

Gene Name	Chromosome	Group with relative activation
PIAS2	18	Recurring tumours
SALL3	18	Recurring tumours
SIGLEC15	18	Recurring tumours
SLC14A1	18	Recurring tumours
SLC14A2	18	Recurring tumours
SMAD7	18	Recurring tumours
SYT4	18	Recurring tumours
TCEB3B	18	Recurring tumours
TCEB3C	18	Recurring tumours
TCEB3CL	18	Recurring tumours
ZNF516	18	Recurring tumours

Supplementary Table 7.S2 – Genes significantly associated with recurrence following treatment with surgery and adjuvant chemotherapy based on analysis of non-squamous lung tumours (n = 52).

Gene name	Chromosome	Group with relative activation
C2orf15	2	Recurring tumours
C2orf29	2	Recurring tumours
CREG2	2	Recurring tumours
EIF5B	2	Recurring tumours
IL18R1	2	Recurring tumours
IL18RAP	2	Recurring tumours
IL1R1	2	Recurring tumours
IL1R2	2	Recurring tumours
IL1RL1	2	Recurring tumours
IL1RL2	2	Recurring tumours
LIPT1	2	Recurring tumours
LYG1	2	Recurring tumours
LYG2	2	Recurring tumours
MAP4K4	2	Recurring tumours
MFSD9	2	Recurring tumours
MITD1	2	Recurring tumours
MRPL30	2	Recurring tumours
NPAS2	2	Recurring tumours
RNF149	2	Recurring tumours
RPL31	2	Recurring tumours
SLC9A2	2	Recurring tumours
SLC9A4	2	Recurring tumours
SP110	2	Recurring tumours
SP140	2	Recurring tumours
SP140L	2	Recurring tumours
TBC1D8	2	Recurring tumours
TMEM182	2	Recurring tumours
TSGA10	2	Recurring tumours
TXNDC9	2	Recurring tumours
SEPT7	7	Recurring tumours
ABCA13	7	Recurring tumours
AMPH	7	Recurring tumours
ANLN	7	Recurring tumours
AQP1	7	Recurring tumours
BBS9	7	Recurring tumours
BMPER	7	Recurring tumours
C7orf10	7	Recurring tumours
C7orf16	7	Recurring tumours
C7orf31	7	Recurring tumours
C7orf36	7	Recurring tumours

Gene name	Chromosome	Group with relative activation
C7orf57	7	Recurring tumours
CAMK2B	7	Recurring tumours
CCDC129	7	Recurring tumours
CHN2	7	Recurring tumours
COBL	7	Recurring tumours
CPVL	7	Recurring tumours
CREB5	7	Recurring tumours
CRHR2	7	Recurring tumours
CYCS	7	Recurring tumours
DPY19L1	7	Recurring tumours
EEPD1	7	Recurring tumours
EGFR	7	Recurring tumours
ELMO1	7	Recurring tumours
EPDR1	7	Recurring tumours
EVX1	7	Recurring tumours
FAM183B	7	Recurring tumours
FIGNL1	7	Recurring tumours
FKBP14	7	Recurring tumours
FKBP9	7	Recurring tumours
FLJ21075	7	Recurring tumours
FLJ22374	7	Recurring tumours
GARS	7	Recurring tumours
GGCT	7	Recurring tumours
GHRHR	7	Recurring tumours
GLI3	7	Recurring tumours
GPR141	7	Recurring tumours
HECW1	7	Recurring tumours
HERPUD2	7	Recurring tumours
HIBADH	7	Recurring tumours
HOXA1	7	Recurring tumours
HOXA10	7	Recurring tumours
HOXA11	7	Recurring tumours
HOXA13	7	Recurring tumours
HOXA2	7	Recurring tumours
HOXA3	7	Recurring tumours
HOXA4	7	Recurring tumours
HOXA5	7	Recurring tumours
HOXA6	7	Recurring tumours
HOXA7	7	Recurring tumours
HOXA9	7	Recurring tumours
HUS1	7	Recurring tumours
IKZF1	7	Recurring tumours
INHBA	7	Recurring tumours
INMT	7	Recurring tumours
JAZF1	7	Recurring tumours
KBTBD2	7	Recurring tumours

Gene name	Chromosome	Group with relative activation
KIAA0241	7	Recurring tumours
KIAA0895	7	Recurring tumours
LOC401335	7	Recurring tumours
LOC402644	7	Recurring tumours
LSM5	7	Recurring tumours
NEUROD6	7	Recurring tumours
NOD1	7	Recurring tumours
NPSR1	7	Recurring tumours
NPVF	7	Recurring tumours
NT5C3	7	Recurring tumours
PDE1C	7	Recurring tumours
PKD1L1	7	Recurring tumours
PLEKHA8	7	Recurring tumours
POU6F2	7	Recurring tumours
RALA	7	Recurring tumours
RP9	7	Recurring tumours
SCRN1	7	Recurring tumours
SEC61G	7	Recurring tumours
SFRP4	7	Recurring tumours
SKAP2	7	Recurring tumours
STARD3NL	7	Recurring tumours
SUNC1	7	Recurring tumours
TARP	7	Recurring tumours
TAX1BP1	7	Recurring tumours
TBX20	7	Recurring tumours
TNS3	7	Recurring tumours
UPP1	7	Recurring tumours
VSTM2A	7	Recurring tumours
VWC2	7	Recurring tumours
WIPF3	7	Recurring tumours
ZNRF2	7	Recurring tumours
C9orf79	9	Recurring tumours
CCRK	9	Recurring tumours
CTSL3	9	Recurring tumours
GSN	9	Recurring tumours
STOM	9	Recurring tumours
DRD5	4	Non-recurring tumours
AGAP11	10	Non-recurring tumours
BMPR1A	10	Non-recurring tumours
C10orf116	10	Non-recurring tumours
GRID1	10	Non-recurring tumours
MMRN2	10	Non-recurring tumours
SNCG	10	Non-recurring tumours

Supplementary Table 7.S3 – Genes significantly associated with recurrence following treatment with surgery and adjuvant chemotherapy based on analysis of squamous lung tumours (n = 29).

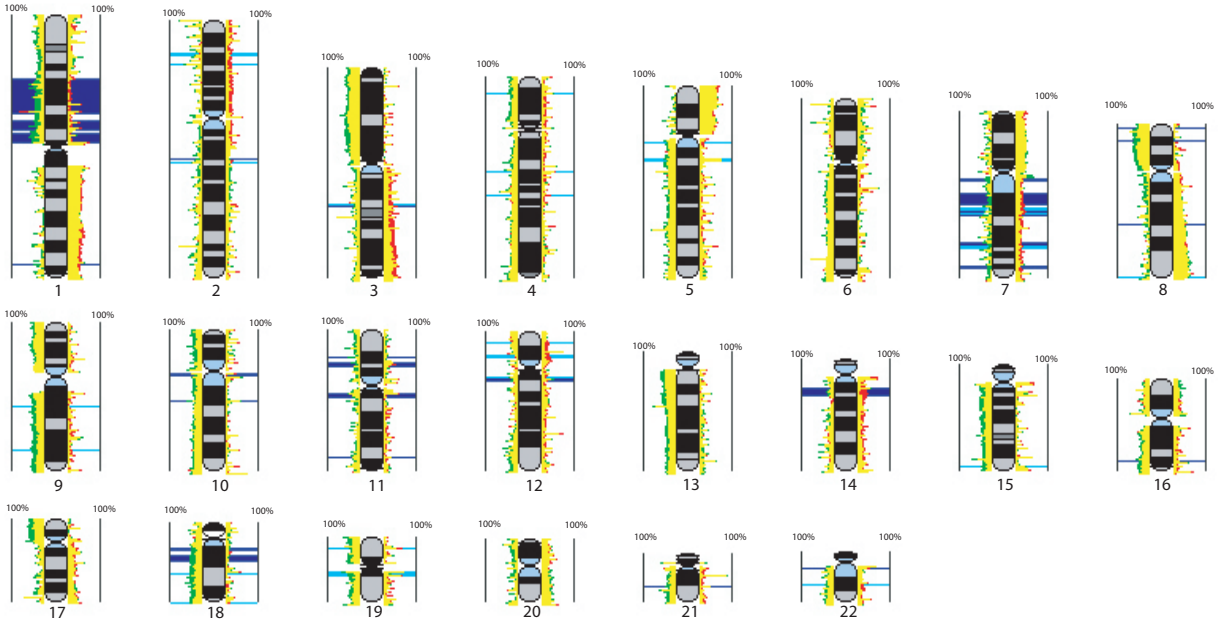
Gene name	Chromosome	Group with relative activation
C1orf53	1	Recurring tumours
CFH	1	Recurring tumours
DENND1B	1	Recurring tumours
TMEM18	2	Recurring tumours
TSSC1	2	Recurring tumours
ROBO1	3	Recurring tumours
ABCG2	4	Recurring tumours
ARHGAP24	4	Recurring tumours
BMPR1B	4	Recurring tumours
DMP1	4	Recurring tumours
FAM13A1	4	Recurring tumours
GRID2	4	Recurring tumours
HERC3	4	Recurring tumours
HERC5	4	Recurring tumours
HERC6	4	Recurring tumours
IBSP	4	Recurring tumours
MAPK10	4	Recurring tumours
MEPE	4	Recurring tumours
MGC48628	4	Recurring tumours
NAP1L5	4	Recurring tumours
PDHA2	4	Recurring tumours
PDLIM5	4	Recurring tumours
PGDS	4	Recurring tumours
PIGY	4	Recurring tumours
PKD2	4	Recurring tumours
PPM1K	4	Recurring tumours
SMARCAD1	4	Recurring tumours
SPP1	4	Recurring tumours
UNC5C	4	Recurring tumours
RASA1	5	Recurring tumours
TMEM167A	5	Recurring tumours
XRCC4	5	Recurring tumours
AIG1	6	Recurring tumours
AKAP7	6	Recurring tumours
ANKRD6	6	Recurring tumours
ARG1	6	Recurring tumours
BACH2	6	Recurring tumours
BVES	6	Recurring tumours
C6orf225	6	Recurring tumours
CASP8AP2	6	Recurring tumours

Gene name	Chromosome	Group with relative activation
CITED2	6	Recurring tumours
ENPP3	6	Recurring tumours
FYN	6	Recurring tumours
GJA10	6	Recurring tumours
GPR126	6	Recurring tumours
HACE1	6	Recurring tumours
HIVEP2	6	Recurring tumours
IRAK1BP1	6	Recurring tumours
LAMA2	6	Recurring tumours
LAMA4	6	Recurring tumours
LIN28B	6	Recurring tumours
LYRM2	6	Recurring tumours
MAP3K7	6	Recurring tumours
MDN1	6	Recurring tumours
MED23	6	Recurring tumours
MOXD1	6	Recurring tumours
OR2A4	6	Recurring tumours
PACRG	6	Recurring tumours
PHIP	6	Recurring tumours
SH3BGRL2	6	Recurring tumours
STX7	6	Recurring tumours
TAAR1	6	Recurring tumours
TAAR2	6	Recurring tumours
TAAR5	6	Recurring tumours
TAAR6	6	Recurring tumours
TAAR8	6	Recurring tumours
TAAR9	6	Recurring tumours
TCF21	6	Recurring tumours
TMEM200A	6	Recurring tumours
TRAF3IP2	6	Recurring tumours
TUBE1	6	Recurring tumours
VNN1	6	Recurring tumours
VNN3	6	Recurring tumours
VTA1	6	Recurring tumours
WISP3	6	Recurring tumours
AMAC1L2	8	Recurring tumours
C8orf74	8	Recurring tumours
DEFB130	8	Recurring tumours
DEFB134	8	Recurring tumours
DEFB136	8	Recurring tumours
DEFB137	8	Recurring tumours
DUB3	8	Recurring tumours
EFHA2	8	Recurring tumours
FAM86B1	8	Recurring tumours
FGF20	8	Recurring tumours
FGL1	8	Recurring tumours

Gene name	Chromosome	Group with relative activation
GSR	8	Recurring tumours
GTF2E2	8	Recurring tumours
MSR1	8	Recurring tumours
MTMR9	8	Recurring tumours
MTUS1	8	Recurring tumours
NAT1	8	Recurring tumours
PDGFRL	8	Recurring tumours
PINX1	8	Recurring tumours
PPP2CB	8	Recurring tumours
PURG	8	Recurring tumours
RBPMS	8	Recurring tumours
SGCZ	8	Recurring tumours
SLC7A2	8	Recurring tumours
SOX7	8	Recurring tumours
TEX15	8	Recurring tumours
TUSC3	8	Recurring tumours
UBXN8	8	Recurring tumours
WRN	8	Recurring tumours
XKR6	8	Recurring tumours
ZDHHC2	8	Recurring tumours
ZNF705D	8	Recurring tumours
ADAMTS20	12	Recurring tumours
BTG1	12	Recurring tumours
CLLU1	12	Recurring tumours
CLLU1OS	12	Recurring tumours
EEA1	12	Recurring tumours
FBXO21	12	Recurring tumours
IRAK4	12	Recurring tumours
LOC338809	12	Recurring tumours
LRRK2	12	Recurring tumours
MRPL42	12	Recurring tumours
NOS1	12	Recurring tumours
NUDT4	12	Recurring tumours
PEBP1	12	Recurring tumours
PLEKHG7	12	Recurring tumours
PUS7L	12	Recurring tumours
SLC2A13	12	Recurring tumours
SOCS2	12	Recurring tumours
SUDS3	12	Recurring tumours
TAOK3	12	Recurring tumours
TESC	12	Recurring tumours
TMCC3	12	Recurring tumours
TMEM117	12	Recurring tumours
TWF1	12	Recurring tumours
UBE2N	12	Recurring tumours
BZRAP1	17	Recurring tumours

Gene name	Chromosome	Group with relative activation
CUEDC1	17	Recurring tumours
DYNLL2	17	Recurring tumours
EPX	17	Recurring tumours
HSF5	17	Recurring tumours
LPO	17	Recurring tumours
MKS1	17	Recurring tumours
MPO	17	Recurring tumours
MRPS23	17	Recurring tumours
OR4D1	17	Recurring tumours
OR4D2	17	Recurring tumours
RNF43	17	Recurring tumours
SFRS1	17	Recurring tumours
SUPT4H1	17	Recurring tumours
VEZF1	17	Recurring tumours
C18orf55	18	Recurring tumours
CYB5A	18	Recurring tumours
FBXO15	18	Recurring tumours
HDHD2	18	Recurring tumours
IER3IP1	18	Recurring tumours
KATNAL2	18	Recurring tumours
TCEB3B	18	Recurring tumours
TCEB3C	18	Recurring tumours
TCEB3CL	18	Recurring tumours

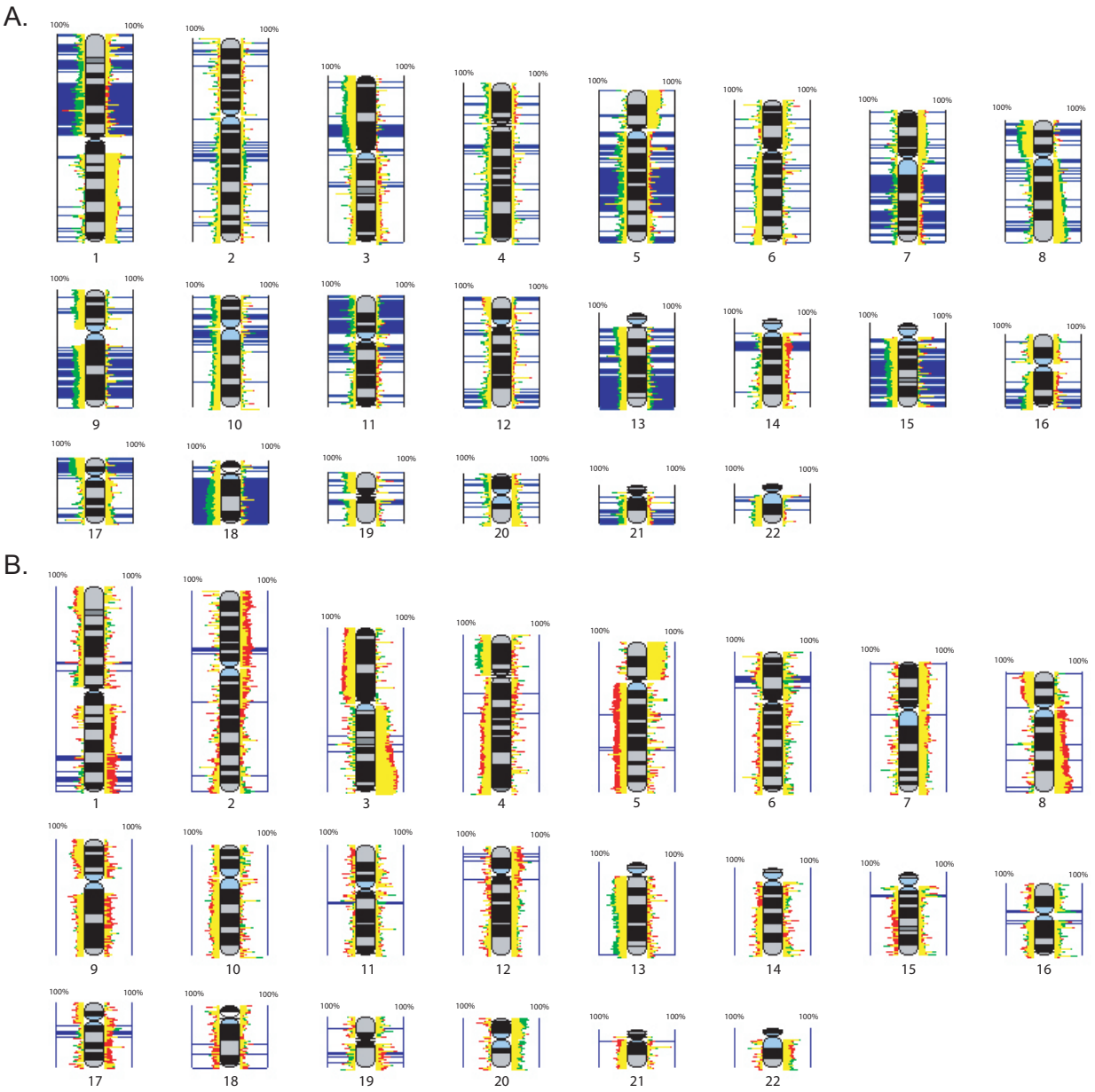
Figure 7.S1



Supplementary Figure 7.S1 – Whole genome frequency of alteration for NSCLC patients treated with surgery alone. Individual tumour genome profiles were generated for each case in a panel of stage IA-IIIA lung tumours (n = 117). Segmental DNA copy number alterations for each case were defined as described in Materials and Methods (Chapter 7). Genome loci were altered to a maximum frequency of 100%. In this image, frequency of DNA gain is defined to the right of each chromosome, while frequency of loss is defined to the left. Red represents "non-recurring" specimens and green represents "recurring" (yellow indicates overlap between these two groups). The lighter blue horizontal lines in the karyogram span genome segments found to be significantly different between response groups and between histological groups (squamous versus non-squamous subtypes) (Fisher's Exact Test, $p < 0.05$). The dark blue horizontal lines span segments significantly associated with response groups alone.

Supplementary Figure 7.S2 – Whole genome frequency of alteration within NSCLC subtypes where patients were treated with surgery alone. Both images are annotated as in Figure 7.S1, with red representing “non-recurring”, green as “recurring”, and yellow indicating an overlap between these groups. Dark blue lines denote regions significantly different between recurrence groups (Fisher’s Exact Test, $p < 0.05$). A. Regions of differential genomic alteration within non-squamous tumours. Specifically, genomic differences were determined for 51 non-recurring versus 31 recurring tumours. B. Regions of differential genomic alteration within squamous tumours. Differences were determined for 26 non-recurring cases versus 9 recurring ones.

Figure 7.S2



Supplementary Figure 7.S3 – Genome profiles comparing chemoresistant and chemosensitive NSCLC cell lines. The karyograms are annotated as in Figure 7.S1 and 7.S2, with red now representing drug sensitive cells and green representing drug resistant cells. A. Results for *cis*-platinum resistance. Genomic data for six resistant lines were compared to data for 11 sensitive lines, as defined by IC_{50} values. B. Analysis of genomic data in the context of vinorelbine resistance. Forty-two drug sensitive cell lines were compared to seven resistant lines.

Figure 7.S3

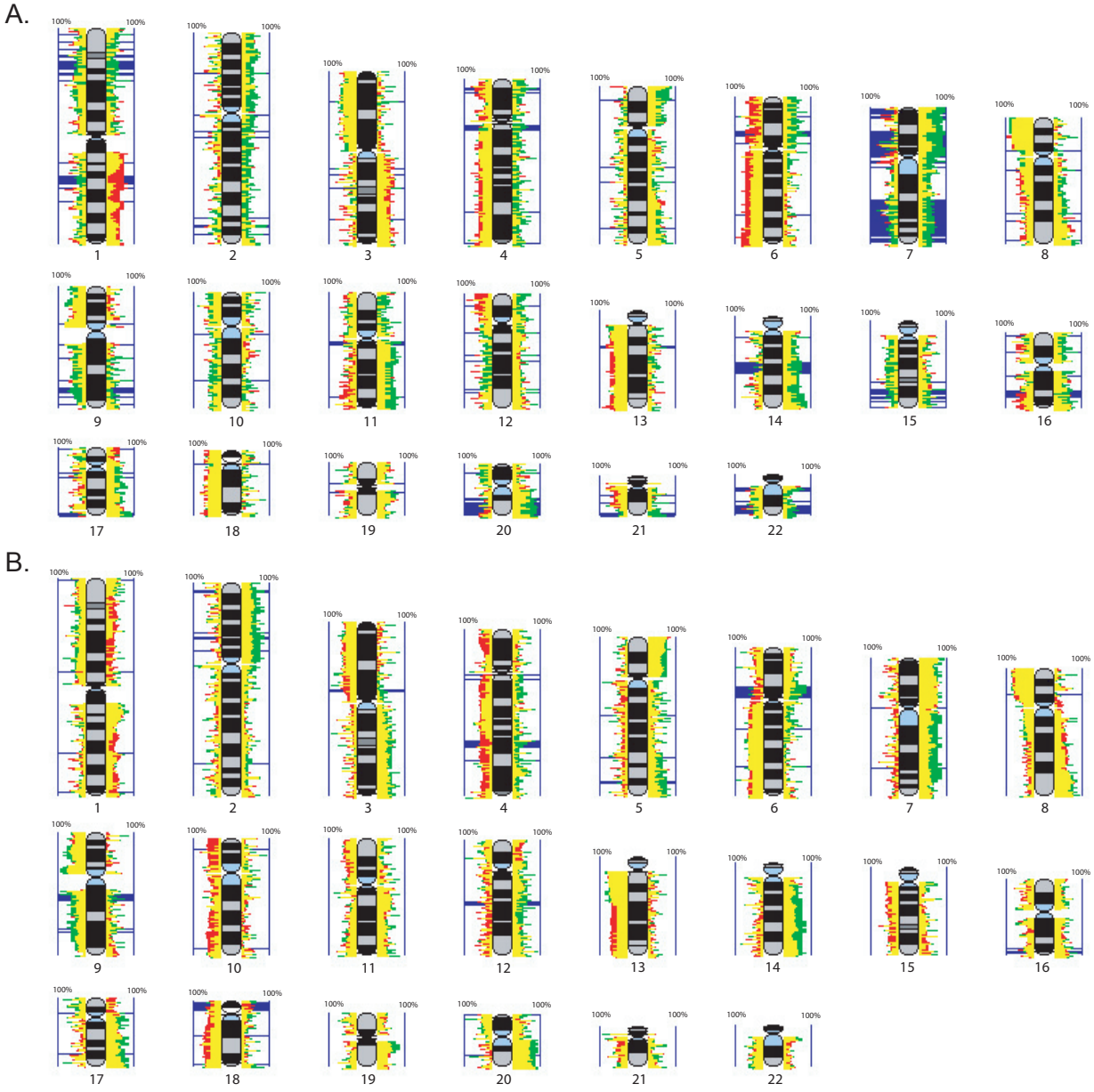
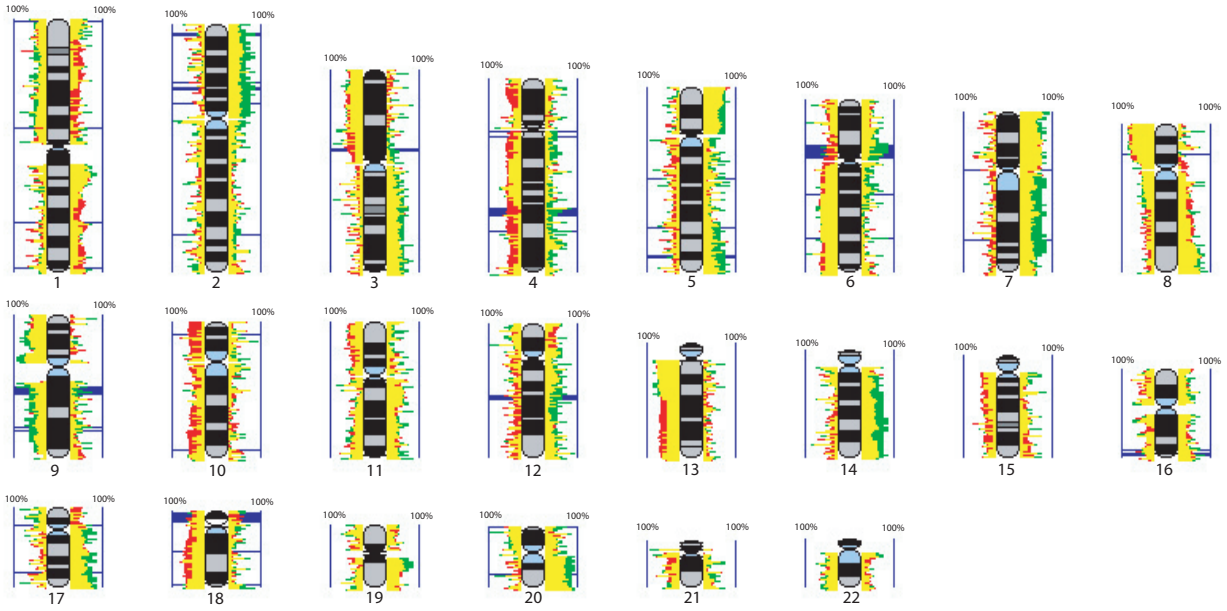

A.
100%
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
B.
100%
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22

Figure 7.S4



Supplementary Figure 7.S4 – Genome profiles comparing vinorelbine-resistant and vinorelbine-sensitive nonsquamous lung cancer cell lines. The karyogram is annotated as in Figure 7.S3. Genomic data for 37 cell lines defined as sensitive to vinorelbine were compared to data for seven lines defined as resistant. This analysis was not repeated in non-squamous lung cancer lines for *cis*-platinum due to the fact that all samples classified as sensitive or resistant in the combined NSCLC analysis were already of non-squamous histology (see Figure 7.S3A).