

**Human and Mouse Genome  
Analysis using Array Comparative  
Genomic Hybridization**

Snijders, Antoine M.

Human and Mouse Genome Analysis using Array Comparative Genomic Hybridization

by Antoine Maria Snijders. - University of California San Francisco  
Utrecht, Utrecht University, 2004.

Proefschrift. - ISBN 90-393-3786-1

Keywords: array comparative genomic hybridization, DNA copy number, genetic instability,  
amplification, cancer, head and neck squamous cell carcinoma

© 2004, A.M. Snijders

Printed by Eindhoven University Press

# **Human and Mouse Genome Analysis using Array Comparative Genomic Hybridization**

Genomische Analyze van Mens en Muis met Array  
'Comparative Genomic Hybridization'  
(met een samenvatting in het Nederlands)

PROEFSCHRIFT

ter verkrijging van de graad van doctor  
aan de Universiteit Utrecht  
op gezag van de Rector Magnificus, Prof. dr. W.H. Gispen,  
ingevolge het besluit van het College voor Promoties  
in het openbaar te verdedigen  
op vrijdag 3 december 2004 des middags te 12:45 uur

door

Antoine Maria Snijders

geboren op 21 juni 1977 te Nunhem

Dit proefschrift is goedgekeurd door de promotoren:

prof.dr. P.J. van Diest  
prof.dr. D.G. Albertson

Copromotoren:

prof.dr. G.A. Meijer  
prof.dr. D. Pinkel

*This research was financially supported by NIH grants CA83040, CA84118, CA94407, CA90421  
and Vysis-Abbott, Inc.*

*To my father and mother*



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Array CGH measures DNA copy number aberrations . . . . .	1
1.2	Constitutional DNA copy number aberrations and polymorphisms . . . . .	2
1.3	Segmental and sequence variations in laboratory mice . . . . .	4
1.4	Tumor profiling by array CGH: instability and selection . . . . .	4
1.5	Tumor profiling by array CGH: studying the genomic relatedness of tumors . . . . .	6
1.6	Cell cycle and checkpoint regulation . . . . .	6
1.7	Aims and Outline . . . . .	7
<b>2</b>	<b>Current Status and Future Prospects of Array Based Comparative Genomic Hybridization</b>	<b>11</b>
2.1	Abstract . . . . .	12
2.2	Introduction . . . . .	12
2.3	Microarray CGH . . . . .	13
2.4	Technological Approaches to Array CGH . . . . .	15
2.5	Applications . . . . .	18
<b>3</b>	<b>Assembly of microarrays for genome-wide measurement of DNA copy number by CGH</b>	<b>21</b>
3.1	Abstract . . . . .	22
3.2	Results and Discussion . . . . .	22
3.3	Acknowledgements . . . . .	24
<b>4</b>	<b>Mapping segmental and sequence variations among laboratory mice using BAC array CGH</b>	<b>25</b>
4.1	Abstract . . . . .	26
4.2	Introduction . . . . .	26
4.3	Results . . . . .	27
4.3.1	Mapping large scale variation in mouse genomes . . . . .	27
4.3.2	Organization of variant loci in the genome . . . . .	27
4.3.3	Small-scale sequence variation between mouse species allows mapping of genomic content in interspecific backcrosses . . . . .	31
4.4	Discussion . . . . .	32
4.5	Materials and Methods . . . . .	35
4.5.1	DNA Extraction . . . . .	35
4.5.2	Preparation of Mouse BAC Arrays . . . . .	36

---

4.5.3	DNA labeling by random priming . . . . .	36
4.5.4	Array hybridization . . . . .	37
4.5.5	Imaging and Analysis . . . . .	37
4.5.6	Statistical methods . . . . .	37
4.5.7	BAC end sequencing ananalysis . . . . .	38
4.5.8	Fluorescent <i>in situ</i> hybridization (FISH) . . . . .	39
4.6	Acknowledgements . . . . .	39
<b>5</b>	<b>Shaping of Tumor and Drug Resistant Genomes by Instability and Selection</b>	<b>41</b>
5.1	Abstract . . . . .	42
5.2	Introduction . . . . .	42
5.3	Results . . . . .	43
5.3.1	Copy number aberrations in MMR deficient and proficient cell lines . . . . .	43
5.3.2	Copy number aberrations in methotrexate resistant MMR deficient and proficient cells . . . . .	44
5.4	Discussion . . . . .	49
5.5	Materials and Methods . . . . .	52
5.5.1	Cells . . . . .	52
5.5.2	Cell culture and selection of resistant cell pools . . . . .	52
5.5.3	DNA isolation . . . . .	53
5.5.4	DNA labeling . . . . .	53
5.5.5	BAC arrays . . . . .	53
5.5.6	Array CGH hybridization . . . . .	54
5.5.7	Imaging and analysis . . . . .	54
5.5.8	Statistical methods . . . . .	54
5.6	Acknowledgements . . . . .	55
<b>6</b>	<b>Genome-Wide Array Based Comparative Genomic Hybridization Reveals Genetic Homogeneity and Frequent Copy Number Increases Encompassing <i>CCNE1</i> in Fallopian Tube Carcinoma</b>	<b>57</b>
6.1	Abstract . . . . .	58
6.2	Introduction . . . . .	58
6.3	Results . . . . .	58
6.4	Acknowledgements . . . . .	63
<b>7</b>	<b>Demonstration of Oral Squamous Cell Carcinoma Second Primary Clonality Using Genome Wide Array Comparative Genomic Hybridization</b>	<b>65</b>
7.1	Introduction . . . . .	66
7.2	Case Reports . . . . .	67
7.2.1	Patient 1 . . . . .	67
7.2.2	Patient 2 . . . . .	68
7.2.3	Patient 3 . . . . .	69
7.3	Results . . . . .	71
7.3.1	Array CGH analysis of tumors from patients with multiple SCC . . . . .	71
7.3.2	Copy number alterations present in multiple tumors from individual patients . . . . .	72
7.4	Discussion . . . . .	73
7.5	Methods . . . . .	74



7.5.1	Isolation of genomic DNA . . . . .	74
7.5.2	Array CGH . . . . .	74
7.6	Acknowledgements . . . . .	74
<b>8</b>	<b>Rare Amplicons Implicate Frequent Misspecification of Cell Fate in Oral Squamous Cell Carcinoma</b>	<b>75</b>
8.1	Abstract . . . . .	76
8.2	Results and Discussion . . . . .	76
8.3	Materials and Methods . . . . .	82
8.3.1	Tumor Samples . . . . .	82
8.3.2	Isolation of DNA . . . . .	82
8.3.3	Array CGH . . . . .	82
8.3.4	Statistical Analysis . . . . .	83
8.3.5	Quantitative RT-PCR . . . . .	83
8.3.6	<i>TP53</i> sequencing . . . . .	83
8.4	Acknowledgements . . . . .	83
<b>A</b>	<b>Supplementary information: Assembly of microarrays for genome-wide measurement of DNA copy number by CGH</b>	<b>85</b>
A.1	Methods . . . . .	85
A.1.1	Specimens . . . . .	85
A.1.2	Genomic clones . . . . .	85
A.1.3	Isolation of BAC/P1 DNA . . . . .	86
A.1.4	Preparation of BAC/P1 DNA representations by ligation-mediated PCR	86
A.1.5	Preparation of DNA spotting solutions . . . . .	87
A.1.6	Array printing . . . . .	87
A.1.7	DNA labeling . . . . .	88
A.1.8	Hybridization . . . . .	88
A.1.9	Imaging and analysis . . . . .	89
A.2	Supplementary Figures and Table . . . . .	90
<b>B</b>	<b>Supplementary information: Shaping of Tumor and Drug Resistant Genomes by Instability and Selection</b>	<b>97</b>
<b>C</b>	<b>Supplementary information: Rare Amplicons Implicate Frequent Misspecification of Cell Fate in Oral Squamous Cell Carcinoma</b>	<b>101</b>
	<b>Summary and discussion</b>	<b>105</b>
	<b>Samenvatting</b>	<b>111</b>
	<b>Bibliography</b>	<b>115</b>
	<b>Acknowledgements</b>	<b>131</b>
	<b>List of Publications</b>	<b>133</b>
	<b>Curriculum Vitae</b>	<b>137</b>



# Chapter 1

## Introduction

### 1.1 Array CGH measures DNA copy number aberrations

Almost all human cancers as well as developmental abnormalities are characterized by the presence of genetic alterations, most of which target a gene or a particular genomic locus resulting in altered gene expression and ultimately an altered phenotype. Understanding which genes or genomic loci are involved in cancer development, progression and maintenance in addition to the characterization of genomic defects in developmental abnormalities, will lead to a better understanding of these complex human diseases as well as identify targets for therapeutic intervention. Different types of genetic alterations include: DNA copy number gains/amplifications, deletions, inversions, translocations, point mutations and epigenetic gene modifications. Over the years, different technological advances have made it possible to detect many of these in a high throughput fashion.

One such technique, comparative genomic hybridization, allows detection of genetic aberrations that result in a change in DNA copy number. In this technology, two genomic DNA pools are differentially labeled using two fluorochromes and hybridized together on a slide containing normal metaphase spreads (Kallioniemi *et al.*, 1992). After hybridization one measures the fluorescence intensities of both fluorochromes along the long axis of each of the chromosomes. Since the hybridization intensity is directly proportional to the copy number, the fluorescence ratio of the two genomic DNA pools is a measure of the DNA copy number difference between the two hybridized samples. Typically one hybridizes one sample with known DNA copy number, such as normal genomic DNA, and one sample with unknown DNA copy number, such as a tumor genomic DNA. Regions of the tumor genome that were present at 2 copies would then give a ratio of 1 ( $2/2$ ), while a single copy gain and loss would result in a ratio of 1.5 ( $3/2$ ) and 0.5 ( $1/2$ ), respectively.

The main advantage of this technology is that one acquires a DNA copy number profile of the entire genome from the sample under investigation. However, typically only DNA copy number losses larger than 10 megabases in size can be detected. For DNA copy number gains this can be less depending on the size and the amplitude. As the human genome project progressed, large sets of mapped clones became available. This, together with the advances that were being made in the field of microarrays, led to the development of array based CGH (array CGH), which was carried out largely within the framework of this PhD thesis (chapters 3 and 4), an initiative of the University of California at San Francisco (UCSF) (Pinkel

*et al.*, 1998; Snijders *et al.*, 2001). The principle of array CGH is identical to original (or chromosome) CGH in that two genomes are hybridized together after which the fluorescence intensities are measured for the two channels and a ratio is calculated which is indicative of the relative DNA copy number difference between the two samples. The major difference, however, is that instead of hybridizing to metaphase chromosomes; one hybridizes to an array of clones mapped on the genome. The resolution is now mainly dependent on the distance between the arrayed clones as well as the size of the clones. For example, an array of 3000 human clones across the genome would result in an average resolution of  $\sim 1$  Mb (Snijders *et al.*, 2001). By using large insert clones, such as BACs (bacteria artificial chromosome), one can obtain an accurate ratio measurement on each clone individually. Also, since most of these clones are placed directly (or indirectly via STS (sequence tagged site) content for example) onto the sequence of the human genome, one can relatively easily assess gene content in aberrant regions. A comprehensive review of the current status of array based CGH can be found in chapter 2.

Given the accuracy and robustness by which one can quantitatively detect copy number aberrations, there are a plethora of applications for this technology. One can map recurrent regions of copy number aberration in a series of tumors and by doing so identify possible driver genes for tumor formation, progression or maintenance (chapters 6 and 8). One can also investigate the origin of metastases and the relatedness of local recurrences and second primaries (chapter 7). Furthermore, many developmental abnormalities are characterized by low-level, sometimes small, DNA copy number gains/losses that can be hard to identify by karyotyping, array CGH has proven useful in this area as well (Klein *et al.*, 2004; Rauen *et al.*, 2002). Recently, array CGH was used to detect DNA copy number aberrations in spontaneous miscarriages, 50% of which result from fetal chromosome aberrations (Schaeffer *et al.*, 2004). In some instances, array CGH can be used to replace FISH (fluorescent *in situ* hybridization) with the advantages that one obtains a relative DNA copy number for as many clones as were arrayed and the fact that one does not need prior knowledge regarding the locus one wants to assay, making array CGH extremely efficient and high-throughput. However, FISH, compared to array CGH, has the capability to measure absolute copy numbers in single cells in stead of a relative DNA copy number from a, possibly heterogeneous, mixture of cells. Finally, from a different point of view, one can start to investigate what the relationship is between underlying genetic defects and the types and numbers of DNA copy number aberrations present in the tumor (chapter 5).

## 1.2 Constitutional DNA copy number aberrations and polymorphisms

As mentioned in the previous section, array CGH is by no means limited to the detection of DNA copy number aberrations in tumors. Other copy number aberrations detected by array CGH include genomic aberrations associated with different genomic constitutional disorders and certain types of polymorphism. Different recombination schemes underlie many genomic disorders. Normally, homologous recombination, whereby maternally and paternally derived homologous chromosomes exchange identical genomic segments without rearranging the genome, occurs during meiosis and ensures genetic variability within a population. Homologous recombination is also involved in DNA repair. Other 'aberrant' forms of recombination include non homologous recombination and non allelic homologous recombination

(Shaw and Lupski, 2004). Non homologous recombination, i.e. recombination between DNA segments on non homologous chromosomes that share a high degree of sequence similarity, can lead to duplications, deletions, inversions as well as (non) reciprocal translocations, all of which may give rise to different genetic syndromes and cancer. Non allelic homologous recombination refers to a recombination process in which highly similar low-copy repeats (LCRs; also known as segmental duplications) recombine during meiosis resulting in gametes that either harbor a duplication or a deletion (when LCR sequences are in the same orientation) or an inversion (when the LCR sequences are inverted) of the genomic segment in between the LCRs. Many genomic disorders are the result of dosage-sensitive genes present within the recombined sequences. For example, a one copy deletion of the alpha-globin gene which is the result of recombination between flanking LCR sequences results in alpha-thalassemia (Lupski, 1998). Most disorders that arose through one of these mechanisms have recurrent breakpoints, namely, the flanking LCR sequences that initiated the recombination event. Genomic segments other than LCRs, such as (peri) centromeric or telomeric sequences have also been associated with recurrent and non recurrent rearrangements in some genomic disorders. For example, the majority of patients with velo-cardio-facial syndrome (VCF or DiGeorge syndrome) carry a small deletion on chromosome 22q11.2, resulting from a recombination event between LCRs on this chromosome. In addition, VCF patients were identified with non recurrent translocation breakpoints which mapped within a specific LCR on 22q11.2 and a telomeric band on the translocation partner chromosome, suggesting not just LCR sequences are involved in the mechanisms underlying the disorders (Spiteri *et al.*, 2003). Array CGH has been used to detect deletions in VCF patients and was 100% concordant with fluorescent *in situ* hybridization (FISH), the current method of choice in diagnosing patients with this syndrome (Rauen *et al.*, in preparation). One of the major advantages of using array CGH as a clinical diagnostic tool for detection of these deletions is the redundancy of making metaphase spreads. Array CGH has also been used to more accurately map deletion breakpoints in patients with Prader-Willi syndrome (Klein *et al.*, 2004) and to identify novel microdeletions and microduplications not easily detected by conventional karyotyping (Visers *et al.*, 2003; Koolen *et al.*, 2004).

Also, array CGH allows the detection of DNA copy number polymorphisms between different individuals. These changes in copy number are distinct from genomic disorders in the absence of a clear phenotypic defect, but might still have clinical consequences over time. For example, the *apolipoprotein (a)* gene codes for a serum protein, which contains a variable number of repeats of a 'kringle-4' motif. Depending on the number of these repeats in the two genomic DNA pools analyzed by array CGH, one can observe an increase, a decrease or no change in DNA copy number. The size of this protein is largely determined by the number of 5.5 kb kringle-4 domains and the plasma concentration was shown to be inversely correlated with the number of kringle-4 domains (Gavish *et al.*, 1989). Elevated plasma levels have been correlated with an increased risk of coronary heart disease, stroke and atherosclerosis. Moreover, the mere presence of polymorphisms, for example, in the form of low-copy repeats (or segmental duplications) might increase the rate of aberrant recombination, some of which might lead to the above mentioned genomic disorders. Another region that exhibits DNA copy number variation between different individuals is located on chromosome 8p23.1 and includes at least three antimicrobial beta-defensin genes. Individuals can have 2 to 12 copies of a repeat unit which is at least 240 kb long. Studies have shown a significant correlation between copy number of *beta-defensin 4* and mRNA expression of this gene (Hollox *et al.*, 2003). Beta-defensins are expressed in many types of epithelia and have antimicrobial

properties. Also, 8p23.1 is often involved in chromosomal rearrangements. Thus it seems reasonable to investigate the presence of DNA copy number polymorphisms, or variant loci, among different individuals or populations.

### 1.3 Segmental and sequence variations in laboratory mice

Studying genomic variation in humans is complicated since the human population is largely outbred. On the other hand, the inbred nature of many laboratory mouse strains and the sequence conservation compared to humans provides an excellent platform for studying the presence of polymorphisms and their contribution to certain phenotypes. Genomic evolution, much like tumor development, is driven by changes brought about at all levels of genome organization. These changes can range from single nucleotide sequence variations to complex genomic rearrangements such as whole genome duplications or polyploidization, duplications, deletions and inversions of parts of the genome. In the presence of selection these changes might lead to increased fitness and eventually speciation. Although polyploidization is believed to have only played a role early on in vertebrate evolution, duplications are thought to have occurred much more recently (Eichler, 2001).

Segmental duplications, which make up approximately 5% of the human genome (Lander, 2001), are defined as a stretch of genomic DNA, whose size can range from 1 to >200 kb, and which are present in more than one location in the genome. They can be defined as either intra (chromosome specific) or inter (present on non-homologous chromosomes) chromosomal and can contain repeat sequences and/or gene specific sequences (Samonte and Eichler, 2002; Thomas *et al.*, 2003; Eichler, 2001; Horvath *et al.*, 2001). While interchromosomal segmental duplications are predominantly located in regions close to the centromere and/or telomere, and contain few functional genes, intrachromosomal segmental duplications are predominantly found in euchromatic regions, and can contain functional genes and/or gene families (Samonte and Eichler, 2002; Thomas *et al.*, 2003). A two step model to explain the occurrence of segmental duplications has been proposed (Samonte and Eichler, 2002). In this model a host locus donates a copy its genomic DNA to an acceptor site. This can occur multiple times, with different hosts donating different parts of genomic DNA to one or more acceptor sites, creating mosaic blocks of duplications. In a second step, these mosaic blocks are duplicated either as a whole or in part to other genomic locations. Recently, the presence of these types of duplications has been described in laboratory mice (Thomas *et al.*, 2003) (chapter 4) and they are likely to underlie some of the phenotypic diversity observed in laboratory mice.

### 1.4 Tumor profiling by array CGH: instability and selection

One of the aforementioned applications of array CGH was to map recurrent regions of copy number aberration in a series of tumors and thereby identify genes that play a role in cancer. Genes whose expression is favorable for tumorigenesis are likely to be increased in copy number, while regions of the genome containing genes detrimental to the tumor are lost. However, cancer is a very complex disease in which expression levels of one gene or pathway can be altered by many different mechanisms. These mechanisms can generally be classified as epi-

genetic, genetic or post-transcriptional. This can lead to, for example, one tumor amplifying a certain gene to increase its expression, while another tumor may over-express the same gene by acquiring a point mutation in a regulatory region rendering it constitutively active. Yet another tumor might alter a gene upstream or downstream of the same or even a different pathway to achieve the same overall goal. The question that comes to mind is why different tumors find different solutions to the same problem? Most tumors are fundamentally different in their genetic background and therefore in their ability to deal with different types of stress. However, the relationship between a tumor's genetic background and the types of genomic aberrations present is not well understood for the majority of solid tumors. Mismatch repair (MMR) deficient colon tumors are perhaps the most well characterized example. These tumors have a defect in one of the mismatch repair genes, typically *MSH2/6* or *MLH1* resulting in a high level of microsatellite instability. As a result, they acquire mutations in genes that are more susceptible to be mutated in a MMR deficient background and those that are found in the tumor are expected to have provided a certain advantage for the cell. In addition, MMR deficient tumors harbor few aberrations at the chromosome level. Another example describing the relationship between underlying genetic defect and the number of copy number aberrations present are *BRCA1* deficient breast tumors. These tumors have been shown to have greater levels of aneuploidy than *BRCA1* proficient breast tumors. *BRCA1* plays a role in many cellular processes including DNA damage signaling and checkpoint activation, suggesting a role for *BRCA1* in maintaining genomic integrity.

In the hereditary forms of colon and breast cancer, patients inherit a mutated allele of a MMR gene (e.g. *MSH2/6* or *MLH1*) or *BRCA1*, respectively. This suggests that in hereditary forms of these cancers, the initial inherited mutation, in a gene whose normal function is to maintain the integrity of the genome, determines what type of genetic instability facilitates progress towards cancer. It should be noted that these initial mutations do not give cells proliferative advantage per se, but rather provide these cells with an increased mutation rate. Positive selection then leads these precancerous cells down a certain path, along which they most likely acquire aberrations (or solutions) that arise fastest and are the best solution to the problem at hand (i.e. growth and survival). It is these cells that will continue to progress towards a true cancer.

Thus it is likely that the types of genomic aberrations a tumor acquires might give a clue as to what the underlying defect in maintenance of genomic instability is and possibly be helpful in tumor drug development or in predicting the response of a tumor to a certain treatment. However, one should consider the possibility that most solid tumors have acquired not just one type of genetic instability but rather several as long as the cells' viability is not compromised, that is, the cost should never outweigh the benefit. Tumor genomes are then likely to be a mosaic of different types of aberrations. By using array CGH one can start to unravel this question at the DNA copy number level, since by using this technique one can distinguish different types and numbers of DNA copy number aberrations. Furthermore, one can accurately define breakpoints and thereby also determine if certain regions of the genome are more prone to rearrange.

## 1.5 Tumor profiling by array CGH: studying the genomic relatedness of tumors

In head and neck squamous cell carcinoma (HNSCC), clinicians often observe patients who present with a single SCC. A second tumor in these patients is most of the time a recurrence of the primary tumor. Other patients present with multiple synchronous tumors and often develop metachronous tumors. The emergence of multiple primary tumors and local recurrences can be explained by several different hypotheses. First, tumors can be independent and arise from different lesions or from a field of (pre)neoplastic but unrelated cells that was exposed to the same carcinogen for a prolonged period of time. This was called "field cancerization" (Tabor *et al.*, 2002) and was first introduced when studying oral squamous cell carcinoma (oral SCC) by Slaughter *et al.* (1953). Another possibility is that multiple tumors arise from a patch or field of clonally related cells. These so called "second field tumors" would then share certain aberrations that pre-existed within the field and likely also have tumor specific aberrations. A third possibility is that tumor cells from one tumor slough off and migrate to a distant site. These tumors would share many of their aberrations, although these tumors might have acquired certain tumor specific aberrations depending on the heterogeneity of the primary tumor and their rate of genomic instability and presence of environmental pressure. Although clinically, for oral SCC, a tumor is labeled a second primary if it emerged >2 cm from the primary site and/or the time in between the two tumors exceeds 3 years, the above discussed hypotheses clearly bring up the possibility that these tumors might be related. Currently, numerous studies on determining whether or not recurrent tumors are clonally related have resulted in both conclusions; multiple tumors are (Carey, 1996) and are not (Ai *et al.*, 2001) clonally related. Most of these studies determine relatedness based on relatively few markers. The use of array CGH can be extremely powerful in this regard, since a rather unique DNA copy number profile of the whole genome can be assessed rather easily. One can compare DNA copy number profiles of multiple tumors from one patient and assess which aberrations are shared among the tumors and exactly, with ~1-2 Mb accuracy, determine and compare DNA copy number transitions (breakpoints). Knowledge about the extent of a possible field and whether two tumors are completely independent, arose from the same field or arose through migration and re-seeding of tumor cells could impact treatment decisions. For example, if one knew the extent of a field of aberrant cells in which a tumor arose, it might in some instances be better to resect the entire field including the tumor instead of resecting only the tumor tissue or it might guide post-surgical treatment such as radiotherapy. However, since the assessment of field cancerization or tumor relatedness involves detailed and specialized molecular analysis, most tumors are treated by a standard protocol, which consists of surgery, radiation therapy in certain cases and clinical follow-up. The use of array CGH in addition to fluorescent *in situ* hybridization will help in the understanding of the multifocality of many cancers and ultimately will benefit the patients.

## 1.6 Cell cycle and checkpoint regulation

The cell cycle, which produces two daughter cells from a single cell, is essential for growth and cell renewal and is a fundamental feature of any life form. The cell cycle is divided into four phases: cells enter S-phase, in which their genome is replicated, after emerging from a gap phase termed G<sub>1</sub>. After DNA replication cells enter a second gap (G<sub>2</sub>) phase after which



cells enter mitosis in which sister chromatids are pulled to opposite poles and two daughter cells are formed. Each daughter cell then enters  $G_1$  or  $G_0$  after terminal differentiation. The time cells spend in  $G_1$ , S and  $G_2$  phase is termed interphase, while mitosis is divided into prophase, metaphase, anaphase and telophase. During prophase the nuclear envelope breaks down and DNA condenses. In metaphase, the condensed and replicated chromosomes assemble on the metaphase plate while in anaphase the sister chromatids are pulled to the oppositely oriented spindle poles. In telophase, two nuclei form and the cell membrane pinches to form two daughter cells.

The cell cycle is a sequential process that needs to be completed from start to finish before a cell can initiate another cell cycle program. It is therefore not surprising that this process is tightly regulated and that errors during the cell cycle need to be properly dealt with, either by slowing or stopping the cell cycle allowing for repair or by inducing apoptosis. These surveillance mechanisms are broadly termed checkpoints. It is important to realize that these mechanisms are not necessary for the cell cycle per se but merely ensure proper passage through the cycle. Different DNA maintenance checkpoints monitor different stages of the cell cycle. For example, DNA damage checkpoints monitor the genome for DNA damage, replication checkpoints ensure proper DNA replication and mitotic checkpoints monitor chromosome segregation and cell division. Lack or defect in parts of these DNA maintenance checkpoints leads to genomic instability since these cells can not respond to DNA damage or replication errors. One can imagine that this may play a role in tumor initiation and pathogenesis. Checkpoint proteins can be divided into three categories. Sensor proteins detect the DNA damage and usually bind the damaged DNA. Transducer proteins are typically protein kinases that are recruited to the sites of DNA damage by the sensor proteins and function as signal amplifiers. Effector proteins are phosphorylated by the transducers and further the signaling cascade into DNA repair and/or growth arrest and apoptosis. One should not think of these pathways as linear but rather as complex networks with the necessary redundancy to maintain a strong response system. (Nyberg *et al.*, 2002; Hanahan and Weinberg, 2000)

Recently, the possible role of checkpoint kinases as targets for therapeutic intervention has been proposed. It has been postulated that checkpoint inhibition might sensitize tumor cells by inhibiting a proper response after DNA damage while at the same time reduce apoptosis levels in normal cells. On the other hand, checkpoint activation might lead to increased repair and thus protection of normal cells, while simultaneously inducing apoptosis levels in tumor cells. (Zhou and Bartek, 2004)

## 1.7 Aims and Outline

As mentioned before, almost all human cancers as well as developmental abnormalities are characterized by the presence of genetic alterations, some of which may give rise to DNA copy number gains and/or losses of genomic material. The initial focus of this thesis was on the development of a robust genome-wide array CGH platform for both human and mouse. After that, the focus shifted towards application based studies in which array CGH was used to further our understanding of the biology of different cancer types.

**Aim 1:** Development of a high-throughput method for generating DNA spotting solutions from small amounts of BAC (Bacterial Artificial Chromosome) DNA for genome-wide detection of DNA copy number variations in human and mouse genomes using comparative

genomic hybridization technology.

**Aim 2:** Do different mouse strains and species contain strain/species specific DNA copy number polymorphisms? Can these variant loci distinguish different mouse strains and species?

**Aim 3:** What is the role of the genetic background or underlying genetic defect in cancer on the types and numbers of DNA copy number aberrations using the arrays developed in aim 1.

**Aim 4:** Identify recurrent regions of DNA copy number gain or loss in clinical samples and study the clonality of synchronously or metachronously developing tumors using the arrays developed in aim 1.

We chose to build on the available array comparative genomic hybridization (array CGH) technology (Pinkel *et al.*, 1998). In this system, arrays were comprised of large-insert BAC (Bacterial Artificial Chromosome) clones to which two differentially labeled genomic DNA pools were hybridized. The use of BACs, which have an average insert size of  $\sim 150\text{-}200$  kb provide sufficient signal intensity after hybridization for the detection of low-level DNA copy number aberrations when present on the array at a concentration of  $\sim 0.5\text{-}1.0$   $\mu\text{g}/\mu\text{l}$ . However, BACs are single copy vectors, which means that the yield of BAC DNA from a bacterial culture is low compared to, for example, a plasmid preparation. This makes the production of a genome-wide human and mouse array, each comprised of 3000 BACs ( $\sim 1$  Mb resolution) a daunting task. Also, whole BAC DNA, when dissolved in a small volume, is viscous. This compromises our ability to reliably deposit small amounts on glass substrates. Ideally one would have a method that would both amplify and reduce the size of the BAC DNA. Therefore, the first aim of this thesis was to develop a method for generating DNA spotting solutions from human and mouse BAC DNA, onto which we could make accurate DNA copy number measurements and which could reliably be deposited on glass slides. After completion of aim 1, the aim shifted to more application based research using the generated arrays. We explored different means to amplify and reduce the size of BAC DNA, while at the same time maximizing the representation of each BAC in the spotting solution. In chapter 3 we describe our most successful approach. We introduced the first human genome-wide array CGH platform which could be used to detect both low-level DNA copy number aberrations including homozygous deletions as well as high level amplifications in cell lines and clinical samples.

In chapter 4 we show that different mouse strains are indeed characterized by the presence of multiple DNA copy number polymorphisms. The majority of these polymorphisms are recurrent among different individuals from one strain and all analyzed strains can be distinguished based on their polymorphism profile. We also introduce a novel approach for determining regions of heterozygosity in different backcross mice using array CGH. This approach is faster than conventional methods for determining regions of heterozygosity.

In cancer, the types of aberrations are probably dependent on the different selection pressures present in the tumor's environment as well as the underlying genetic instability which may promote certain types and numbers of aberrations in certain genomic locations to occur at a higher frequency. Different tumors are characterized by the presence of different types and numbers of DNA copy number aberrations. However, the relationship between underlying

genetic defect and the actual aberrations present in the tumor is poorly understood, at least for most solid tumors. Although currently most tumors are treated based on their organ of origin and histo-pathological findings, it is likely that once we can also classify tumors based on their underlying genetic defects one can more accurately tailor treatment. In order to study the relationship between underlying genetic defect and types and numbers of DNA copy number aberrations one needs a robust system that can detect a wide spectrum of aberrations, ranging from homozygous deletions and low-level aberrations to high-level amplifications, at a resolution that is high enough to distinguish the different types of aberrations. In chapter 5 we explore the presence of different types and numbers of copy number aberrations in mismatch repair (MMR) deficient cell lines when compared to MMR proficient cell lines. We also discuss the role of the genetic background on the types and numbers of DNA copy number aberrations by using a cell line model system in which we selected for methotrexate (MTX) resistance in cells whose genetic backgrounds had been altered. After selection, the types and numbers of copy number aberrations were analyzed by array CGH. This study showed that different aberrations were associated with MMR status.

In chapter 6 we describe an array CGH study of 14 Fallopian Tube carcinoma. Genes that map in recurrent regions of gain or loss are discussed. In particular, we propose an important role for *CCNE1* in Fallopian Tube carcinoma. Also the overall observed homogeneity between the different tumors is addressed.

In oral cancer as well as other cancers that arise in the head and neck region, clinicians often observe patients with synchronous or metachronous tumors. Some of these tumors develop years, sometimes decades apart at anatomically distant sites within the oral cavity. We addressed the issue of whether these second primary tumors are in some way clonally related to the index tumor. The use of array CGH is helpful in this regard since one can assess a genome-wide view of the DNA copy number aberrations present in multiple tumors of the same patient and combined with the accuracy at which copy number transitions can be defined one can infer whether tumors are related. In chapter 7 we performed array CGH on three patients, all of whom presented with multiple tumors at various times. We clearly show that these tumors have a clonal origin based on the presence of several 'signature' aberrations shared by tumors from one patient.

Chapter 8 describes the use of genome-wide array CGH to identify recurring regions of DNA copy number amplification in oral squamous cell carcinoma. We focussed on those amplifications which size was less than 3 Mb and performed expression analysis on possible candidate genes to identify driver genes for these amplifications. We found genes involved in integrin signaling, survival, adhesion and migration, and, interestingly, Hedgehog and Notch pathway genes to be amplified and over-expressed. The possible roles of these pathways, which are possible targets for therapeutic intervention, in oral SCC are discussed.



## Chapter 2

# Current Status and Future Prospects of Array Based Comparative Genomic Hybridization

Antoine M. Snijders <sup>1,2</sup>, Daniel Pinkel <sup>2,3</sup> and Donna G. Albertson <sup>1,2,3</sup>

<sup>1</sup> Cancer Research Institute, University of California San Francisco, San Francisco CA, USA

<sup>2</sup> University of California San Francisco Comprehensive Cancer Center, San Francisco, San Francisco CA, USA

<sup>3</sup> Department of Laboratory Medicine, University of California San Francisco, San Francisco CA, USA

## 2.1 Abstract

The majority of human cancers as well as many developmental abnormalities harbor chromosomal imbalances, many of which result in the gain and/or loss of genomic material. Conventional comparative genomic hybridization (CGH) has been used extensively to map DNA copy number changes to chromosomal positions. The introduction of microarray CGH provided a powerful tool to precisely detect and quantify genomic aberrations and map these directly onto the sequence of the human genome. In the past several years, several different approaches towards array based CGH have been undertaken. This paper reviews these approaches and presents some of the recently developed applications of this new technology in both research and clinical settings.

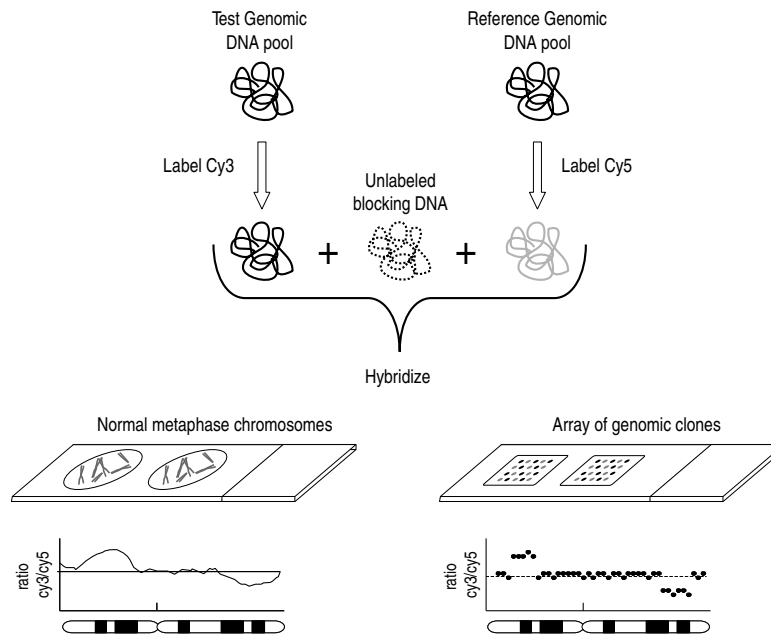
## 2.2 Introduction

Many human cancers as well as many developmental abnormalities are characterized by the presence of genomic DNA copy number changes. In cancer, genes detrimental to tumor growth (e.g. tumor suppressor genes) are likely to be contained in regions of decreased DNA copy number; while genes important for tumor growth or development (e.g. oncogenes) are likely to be contained in regions of increased DNA copy number. Over the past decade, comparative genomic hybridization (CGH) has been used extensively to map DNA copy number changes to a chromosomal position. CGH was first described in 1992 by Kallioniemi *et al.* (1992). In this technology, differentially labeled and denatured genomic DNA pools are simultaneously hybridized to immobilized and denatured chromosome metaphase spreads (Figure 2.1).

After hybridization, the fluorescence ratio is measured along the axis of each chromosome, creating a genome-wide profile of the relative copy number in the DNA pools. Usually, test genomic DNA (e.g. tumor genomic DNA) labeled with one fluorochrome and reference genomic DNA (e.g. normal genomic DNA) labeled with a different fluorochrome are simultaneously hybridized in the presence of high concentrations of unlabeled Cot-1 DNA which blocks repetitive DNA sequences.

The main advantage of CGH is that the whole genome can be interrogated for DNA copy number in a single experiment. After analyzing large numbers of tumors, recurrent regions of copy number alteration can be defined thereby identifying the region(s) containing genes involved in cancer development. One main disadvantage of chromosome CGH is the limited resolution at which DNA copy number aberrations can be detected. Typically, the limit for detection of DNA copy number loss is approximately 10 Mb, but it can be less for copy number gains depending on the size of the involved region and level of copy number increase. Application of statistical methods that correct for variation along the CGH profile can also provide improved resolution (Kirchhoff *et al.*, 1999; Yu *et al.*, 1997).

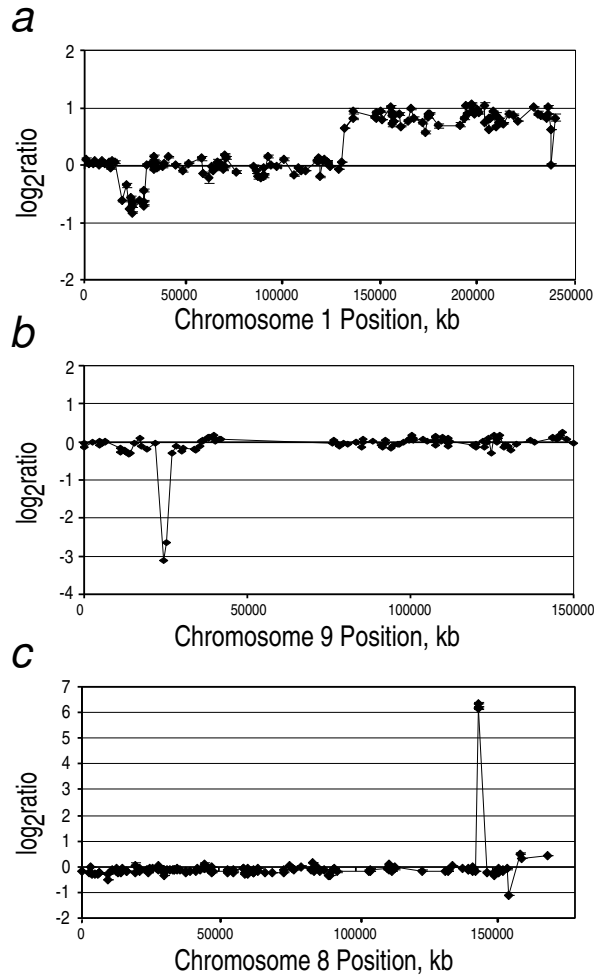
The development of microarray technology opened a way to overcome many of the limitations of CGH, because the metaphase spreads could be replaced by mapped DNA clones, thereby providing a means to link aberrations directly to the genome sequence. Several approaches have been undertaken to implement array-based CGH. The different approaches can be categorized based on the properties of the DNA deposited on the substrate and on the properties of the genomes to be interrogated by the arrays. This review will summarize different approaches towards microarray CGH and their advantages and disadvantages.



**Figure 2.1:** Overview of conventional and array based comparative genomic hybridization (CGH). Test genomic DNA is labeled with one fluorochrome (e.g. Cy3) and reference genomic DNA is labeled with another fluorochrome (e.g. Cy5). Both pools are mixed together with unlabeled blocking DNA, which blocks repetitive sequences in the genome. After denaturation, this mixture is hybridized to normal metaphase spreads, in the case of conventional CGH or to an array of genomic clones, in the case of array based CGH. After hybridization, the fluorescence ratio of the two fluorochromes is measured along the axis of each chromosome in conventional CGH and for each arrayed clone in array based CGH.

### 2.3 Microarray CGH

Microarrays are usually created by high-precision robotics, which deposit and immobilize small amounts of DNA (spots) closely spaced on a substrate (usually a glass surface) in a regular and/or ordered fashion. Typically for microarray CGH, one simultaneously hybridizes differentially labeled genomic DNA pools to each array. The DNAs are labeled in vitro by random priming to incorporate fluorescently labeled nucleotides, usually Cy3 and Cy5. After hybridization, digital images of each of the hybridized fluorochromes can be obtained using various types of imaging systems. Software is used to segment the spots of the hybridized array and calculate a background corrected  $\log_2$  ratio of the hybridized fluorochromes as well as other parameters that assess data quality. The  $\log_2$  ratio would then be a reflection of the relative DNA copy number between the two hybridized specimens for a certain arrayed element. Data are typically normalized so that the median  $\log_2$  ratio = 0. The expected  $\log_2$  ratio of a single copy gain in a diploid genome, when hybridized versus normal genomic DNA, would be +0.58, while for a single copy loss this would be -1. In practice,  $\log_2$  ratio can range from -4 for homozygous deletions to  $\log_2$  ratio >6 for high-level amplifications (Figure 2.2) (Pinkel *et al.*, 1998; Snijders *et al.*, 2001).



**Figure 2.2:** Normalized DNA copy number ratios of clones on chromosomes 1 (a), 9 (b) and 8 (c) in three different cell lines (5, 26). Clones are processed as described previously (Snijders *et al.*, 2001). For each plot, the clones are sorted from the p-arm to the q-arm based on genomic position according to the draft sequence of the human genome (UCSC Golden Path August 2001 freeze, <http://genome.ucsc.edu/>). The different types of DNA copy number aberrations detected by array CGH are a single copy deletion and gain of portions of chromosome 1 (a), a homozygous deletion on chromosome 9 (b) and a high-level amplification on chromosome 8 (c).

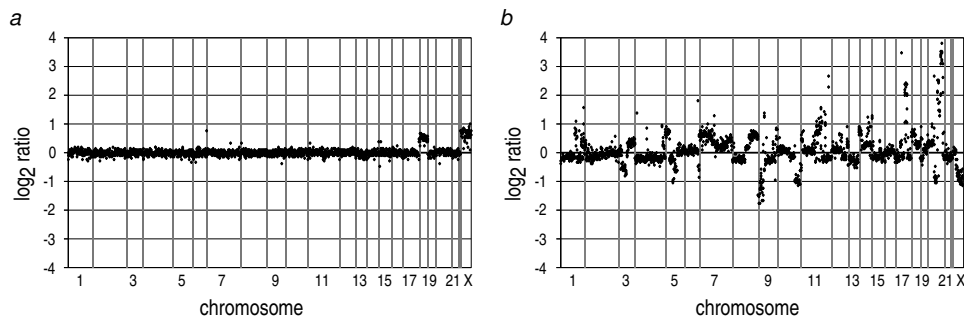
The genomic resolution at which these aberrations can be detected is determined by three major factors; the genomic distance between the arrayed DNA elements, the size of the elements and if ratios are obtained by averaging data from neighboring clones, then the number (and spacing) of the clones used will also affect the resolution. For example, Albertson *et al.* (2000) showed that by arraying overlapping large insert clones, aberrations could be mapped



with genomic resolution ( $\sim 50$  kb) that was less than the size of the clones ( $\sim 150$  kb). In order for microarray CGH to be clinically useful as a diagnostic/prognostic tool, one needs a robust system with the sensitivity to quantitatively detect a wide spectrum of copy number changes, ranging from homozygous deletions and low-level DNA copy number aberrations to high level amplifications. To achieve this level of sensitivity one needs a high signal to noise ratio resulting from high specific hybridization and low background. The hybridization signal intensity depends on many factors, but one critical one is the complexity (size) of the genome being analyzed compared to the complexity (size) of the DNA sequences in the array spots. In general, for a given genome, the signal intensity increases as the complexity of the DNA in the spots is increased. Also, the properties of the DNA specimens interrogated by the array will affect the ratios. The purity of the DNA appears to be particularly important as it affects the ability to label both the test and reference genome equivalently.

## 2.4 Technological Approaches to Array CGH

In the early implementations of array CGH, arrays were comprised of large-insert genomic DNA clones such as BACs (Bacterial Artificial Chromosome) to which differentially labeled genomic DNA pools were hybridized in the presence of Cot-1 DNA (Pinkel *et al.*, 1998; Solinas-Toldo *et al.*, 1997). These large clones containing  $\sim 150$  kb of genomic DNA inserts provided sufficient signal intensity to quantitatively detect single copy number changes as well as homozygous deletions and high-level amplifications. By using BACs that contain an STS and/or have been end or fully sequenced, one can map copy number changes directly onto the sequence of the human genome (Figure 2.3).



**Figure 2.3:** Whole genome array CGH profiles of two cell lines compared to normal reference DNA. The normalized  $\log_2$  ratio is plotted for all clones in genomic order starting at chromosome 1p and ending at chromosome Xq. Clones are processed as described previously (5). Vertical lines indicate chromosome boundaries. **a.** Cell line GM00143 shows a single copy gain of chromosome 18. Note the increase in copy number on the X chromosome due to the sex mismatch between the two hybridized genomic DNA specimens. **b.** Breast cancer cell line BT474 shows multiple gains and losses as well as a high level amplification on chromosome 17 and 20. Note the different populations of clones at varying low-level copy numbers due to aneuploid DNA content.

Sets of STS containing, cytogenetically mapped BAC clones are currently available (Cheung *et al.*, 2001). Since each arrayed element gives reliable data regarding DNA copy number,

BAC arrays can be used to accurately and quantitatively map amplicons, which facilitate the identification of oncogenes (Albertson *et al.*, 2000). Also, because of its accuracy in detecting low-level copy number aberrations, microarray CGH on arrays of BACs can be used in many medical genetic studies (Rauen *et al.*, 2002). However, disadvantages of this approach are the fact that BACs are single copy vectors and the yield of BAC DNA from a standard BAC preparation is low compared to standard plasmid preparations. In addition, accurately depositing sufficient quantities of whole BAC DNA can be problematic due to viscosity of the printing solutions caused by the high-molecular weight DNA. Reducing the size of the BACs by sonication (Pinkel *et al.*, 1998; Bruder *et al.*, 2001; Wessendorf *et al.*, 2002) significantly reduces viscosity and aides in reliably depositing BAC DNA on the substrate. Also, chemical modification of BAC DNA prior to preparation of DNA spotting solution has been described (Cai *et al.*, 2002).

Although the treatments described above overcome difficulties with printing whole BAC DNA, issues regarding production of sufficient quantities of BAC DNA remain. Several PCR based methods that produce a representation of the BAC have been developed to overcome this problem. The advantage of using PCR based methods is the fact that large quantities of DNA can be obtained from a small amount of starting material. However, the main concern is how democratically the method amplifies sequences within the BAC. Ideally, all sequences should be present in equal concentration in the DNA spotting solutions. If the BAC sequence is not well represented in the amplified spotting solutions, it will not provide sufficient signal intensity after hybridization for the accurate detection of single copy number changes in heterogeneous samples. However, because of the inherent nature of any PCR based method some loss of sequence representation will occur. Thus PCR-based methods must be deployed that aim at maximizing the representation of BAC DNA sequence.

Snijders *et al.* (2001) used ligation-mediated PCR (Klein *et al.*, 1999) to amplify BAC DNA. The DNA is cut with a frequently cutting restriction enzyme after which primers are ligated and the ligated fragments are amplified using these primers. A subsequent second PCR using a fraction of the first ligation-mediated PCR is performed, and this DNA is used for spotting. From one ligation-mediated PCR multiple second PCRs can be performed providing large quantities of printing solutions. Because of the use of a frequently cutting restriction enzyme and a high-fidelity DNA polymerase, the BAC sequence appears to be largely maintained and virtually identical DNA copy number ratios can be obtained by hybridizing to the ligation-mediated PCR representation or the whole BAC DNA. This technology has several advantages. The BAC sequence is well represented in the spotting solution and only small amounts of BAC DNA are required initially for the ligation-mediated PCR, which can then be used as template for 40-50 second PCRs, to make printing solutions. However, BAC sequence will be underrepresented in the spotting solution if a certain sequence is sparse in restriction enzyme recognition sites such that the distance between two adjacent sites is greater than the maximum length that can be spanned by the high-fidelity DNA polymerase. The use of a frequent cutting enzyme with a 4-base recognition site helps to reduce this problem.

Other methods use degenerate oligonucleotide primer PCR (DOP-PCR; Telenius *et al.* (1992)) to amplify BAC DNA (Hodgson *et al.*, 2001; Veltman *et al.*, 2002; Fiegler *et al.*, 2003). As originally described, DOP-PCR uses a mixture of primers each comprised of defined 3' and 5' sequences that flank a random hexanucleotide sequence. Low temperature annealing cycles are used initially to promote annealing at multiple sites by the primer mixture. However, there may be a bias toward amplification of sequences corresponding to the 3' most hexanucleotide of the primer that are close together and in opposite direction. To increase the

level of representation Fiegler *et al.* (2003) designed three DOP primers. The 5' sequences of these primers are identical to the 6MW DOP primer first described by Telenius *et al.* (1992). However, the 3' sequences were designed such that human DNA would be amplified preferentially compared to *E. coli* DNA, a known contaminant of DNA preparations from large insert clones. Three separate amplification reactions are carried out. A second PCR is performed using the first DOP-PCRs as template, but in this reaction amplification is primed with an amine-linked primer in which the 3' end matches the 5' ends of the first round DOP primers. Arrays produced by this approach provided reliable detection of low-level DNA copy number aberrations in a human cell line. Based on these promising first results, one expects that the method will prove to be robust and to have high dynamic range when applied to clinical samples or specimens containing homozygous deletions and/or high level amplifications.

A third approach to perform CGH analysis uses microarrays comprised of mapped cDNA clones (Pollack *et al.*, 1999). Microarrays containing thousands of cDNAs are widely available and have been used extensively to analyze gene expression levels at the mRNA level, by simultaneously hybridizing differentially labeled mRNA pools. The use of cDNA arrays for CGH is appealing because of their widespread availability and because both DNA copy number and expression analysis of one sample can be performed in parallel on the same arrayed elements. Thus, one can readily compare DNA copy number and mRNA expression levels (Kauraniemi *et al.*, 2001; Hyman *et al.*, 2002). However, performing array CGH analysis on cDNA microarrays poses technical challenges for reliable detection of single copy number changes in genomes of mammalian complexity. cDNA clones only represent on average 0.5 - 2 kb of intron lacking genomic sequence, resulting in greatly diminished signal intensities compared to hybridization to BACs. As discussed earlier, a high ratio of signal to noise is necessary to quantitatively and reliably detect low-level DNA copy changes on individual array elements and thus there is generally greater clone-to-clone ratio variation on cDNA arrays. To reduce the noise level when hybridizing genomic DNA to cDNA clones, a moving average of ratios of several adjacent cDNA clones is used (18). This procedure will reduce resolution but does not severely affect the resolution when arrays containing large numbers of cDNAs are used. In smaller genomes, such as yeast, cDNA and oligonucleotide arrays have provided useful data (Winzeler *et al.*, 1999).

Amplification of fluorescent signals is often used in cytogenetic applications and thus might also be useful to augment the low signals obtained when CGH is carried out on cDNA arrays. Heiskanen *et al.* (2000) described tyramide-based amplification of signals obtained after hybridizing genomic DNA to cDNA microarrays. However, in this work, only DNA amplifications greater than 5-fold could reliably be detected, and the method did not allow two different DNAs to be distinguished simultaneously, so that it was necessary to hybridize the test and reference samples to two different microarrays. Thus, the utility of signal amplification methodologies is yet to be demonstrated for array CGH.

Compared to BAC arrays, cDNA microarrays appear to be an attractive format for array CGH, because there is widespread access to arrays containing large numbers of mapped cDNA clones, cDNAs are easier to propagate for preparation of spotting DNA and higher resolution mapping might be possible if reliable measurements were possible on single clones. However, robustness and measurement capability for low-level DNA copy number changes appear to vary considerably between laboratories. Thus, further development of this approach and an understanding of the factors important for success seem necessary before cDNA arrays can be expected to provide the majority of practitioners with reliable detection of low-level DNA

copy number changes, particularly when analyzing non-homogeneous or polyploid clinical samples.

Several CGH approaches have been developed that employ reduced-complexity representation of the genomic DNAs as well as the array elements. Geschwind *et al.* (1998) used a PCR based method in which arrays are comprised of inter-*alu* PCR amplified and cloned fragments of genomic clones, such as cosmids or BACs. Arrays are then hybridized with inter-*alu* PCR amplified and labeled pools of specimen genomic DNA. Since the arrays are comprised of cloned DNA fragments, large quantities can relatively easily be produced by PCR. The use of inter-*alu* PCR for the specimen genomic DNA allows one to use small amounts of genomic DNA, which is important in some applications.

Alternatively, Lucito *et al.* (2000) and Lucito and Wigler (2002) used arrays comprised of cloned human DNA restriction fragments and/or long oligonucleotides hybridized with reduced complexity representations of test and reference genomic DNA. The representations are prepared by digesting the genomic DNA with a restriction enzyme after which primers are annealed, ligated and then amplified by PCR using the primer sequences. Since amplification of smaller fragments is favored in the reaction, the resulting amplified DNA is less complex than the starting sample. The PCR products are then labeled by random priming and processed for hybridization. The arrays are made from individual sequences contained in the genomic representation. These sequences are either isolated and cloned, allowing preparation of DNA for array spots, or they are synthesized. By reducing the complexity of both genomic DNA and arrayed DNA, one can in principle improve the signal to noise ratio. However, in published applications of this technique, ratios were still obtained by averaging data from several spots. The main advantages of this approach are the requirement for only small amounts of input genomic DNA and the fact that large quantities of DNA spotting solutions can be obtained relatively easily. However, the main drawback is the difficulty with which low-level DNA copy number aberrations can be detected. One should as well keep in mind that the use of PCR to amplify pools of genomic DNA for hybridization might result in differential amplification of sequences in the test and reference genomic pools and thus affect ratios.

## 2.5 Applications

Genome-wide array CGH (Snijders *et al.*, 2001) can be used to map recurrent regions of copy number aberrations when analyzing large numbers of tumor samples. Subsequently, assembly of arrays with high-density clone coverage across a region allows further definition of recurrent aberrations. For example, Albertson *et al.* (2000) used array CGH to quantitatively and precisely map amplicon structures in tumors, which aided in the identification of oncogenes. Also, because of the precision and sensitivity, the use of array CGH allows one not only to classify tumors based on their types and numbers of DNA copy number aberrations, it also allows one to ask specific questions about the underlying genetic defects that permit the selection for these aberrations (Snijders *et al.*, 2003). Moreover, because many developmental abnormalities are characterized by the presence of low-level genomic imbalances, many of which result in DNA copy number alterations, array CGH can be used to detect the size and location of these aberrations and correlate these aberrations with the clinically observed phenotype (Rauen *et al.*, 2002; Bruder *et al.*, 2001). Veltman *et al.* (2002) described the use of array CGH for the detection of genomic aberrations in (sub)-telomeric regions, often deleted in mentally retarded patients. Currently, these aberrations are detected in the clinical

laboratory by fluorescent in situ hybridization assays with individual telomere probes. In the future, the throughput of such tests may be increased by applying array CGH to assay all telomeres simultaneously.

The array CGH format has also been applied together with other assays to profile epigenetic changes. For example, Zardo *et al.* (2002) used array CGH in combination with restriction landmark genomic scanning to assess both DNA copy number and methylation status in order to identify genes inactivated by single copy deletion and methylation of the remaining copy. In an alternative approach, Shi *et al.* (2002) assembled arrays of CpG island clones and then hybridized bisulfite treated tumor and normal DNA specimens to the arrays in order to reveal differences in the methylation status at these loci. Other studies have used array CGH to map protein-binding sites in the genome (Horak *et al.*, 2002; Iyer *et al.*, 2001). In this technology genomic DNA is immunoprecipitated using specific antibodies for DNA binding proteins. By competitively hybridizing two differentially labeled immunoprecipitated genomic DNA pools, one can identify protein-DNA binding regions in the genome and compare patterns of protein-DNA binding sites in for example tumor and normal DNA.

In conclusion, array CGH currently provides a reliable platform for the detection of DNA copy number changes for a wide variety of applications. More and more new applications are to be expected in both research and clinical settings.



## Chapter 3

# Assembly of microarrays for genome-wide measurement of DNA copy number by CGH

Antoine M. Snijders<sup>1,2</sup>, Norma Nowak<sup>4</sup>, Richard Seagraves<sup>1</sup>, Stephanie Blackwood<sup>1,2</sup>, Nils Brown<sup>1</sup>, Jeffrey Conroy<sup>4</sup>, Greg Hamilton<sup>1</sup>, Anna Katherine Hindle<sup>1,2</sup>, Bing Huey<sup>1</sup>, Karen Kimura<sup>1</sup>, Sindy Law<sup>1,2</sup>, Ken Myambo<sup>1</sup>, Joel Palmer<sup>1,2</sup>, Bauke Ylstra<sup>1,2</sup>, Jingzhu Pearl Yue<sup>1</sup>, Joe W. Gray<sup>1,3</sup>, Ajay N. Jain<sup>1,2,3</sup>, Daniel Pinkel<sup>1,3</sup> and Donna G. Albertson<sup>1,2,3</sup>

<sup>1</sup> Comprehensive Cancer Center,

<sup>2</sup> Cancer Research Institute and

<sup>3</sup> Department of Laboratory Medicine, University of California San Francisco, San Francisco, CA 94143-0808 and

<sup>4</sup> Roswell Park Cancer Institute, Elm and Carlton Streets, Buffalo, NY 14263

### 3.1 Abstract

We have assembled arrays of ~2400 BAC clones for measurement of DNA copy number across the human genome. The arrays provide precise measurement (S.D. of  $\log_2$ ratios = 0.05-0.10) in cell lines and clinical material, so that we can reliably detect and quantify high level amplifications and single copy alterations in diploid, polyploid and heterogeneous backgrounds.

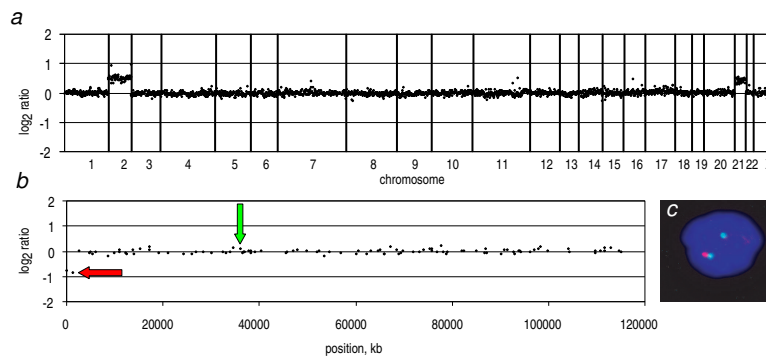
### 3.2 Results and Discussion

Microarray-based comparative genomic hybridization (array CGH) provides a means to quantitatively measure DNA copy number aberrations and to map them directly onto genome sequence. Since arrays comprised of large insert genomic clones such as BACs provide reliable copy number measurements on individual clones (Pinkel *et al.*, 1998; Albertson *et al.*, 2000), they have potential utility for research and clinical applications in medical genetics and cancer. However, preparation and spotting of BAC DNA is problematic. BACs are single copy vectors so the yield of DNA from cultures is low compared to plasmids, and spotting high molecular weight DNA at sufficient concentration to obtain good signal to noise in the hybridizations may be difficult. To overcome these problems, we used a ligation-mediated PCR method (Klein *et al.*, 1999) to generate representations of human and mouse BAC DNAs for spotting on arrays. We produced sufficient spotting solution (0.8  $\mu\text{g}/\mu\text{l}$  DNA in 20% DMSO) from 1 ng of BAC DNA to make tens of thousands of arrays (for detailed Methods, see Appendix A section A.1). The ratios we measured on arrays comprised of BAC representations are essentially identical to ratios previously reported on DNA from the same BACs (Pinkel *et al.*, 1998). Independently prepared DNA representations yield highly reproducible data (average variation of the linear ratios on individual clones from two independent preparations, 6.6%). For copy number assessment across the human genome, we printed 2460 BAC and P1 clones in triplicate (~7500 elements) in a 12 mm square (HumArray 1.14, Appendix A Figure A.1 and Web Table A). Each clone contains at least one STS, allowing linkage to the genome sequence. Cytogenetic mapping indicated that 2298 of the arrayed clones are single copy (Cheung *et al.*, 2001; Knight *et al.*, 2000), and therefore these arrays provide average resolution of ~1.4 Mb across the genome. We have also assembled an array of ~1300 clones for the mouse, which will be reported elsewhere. With the human arrays, we have obtained highly reproducible measurements over a wide dynamic range in cancer cell lines (Appendix A Figure A.5 and Web Tables B and C for analyses of COLO320, HCT116, HT29, MDA-MB-231, MDA-MB-453, MPE600, SW837 and T47D). These copy number alterations ranged from homozygous deletions ( $\log_2$ ratio < -2, HCT116 chromosome 16) to very high level amplifications ( $\log_2$ ratio > 6, amplification of CMYC, COLO320). We also obtained nearly identical ratios (average S.D. of the  $\log_2$ ratio = 0.08) in three replicate hybridizations with BT474 cell line DNA, two labeled by random priming and one by nick translation, using an array of 1777 clones (HumArray 1.11, Appendix A Figure A.2 and Web Table D).

To test our ability to measure single copy changes (i.e. trisomies and monosomies), which is critical for applications involving medical genetics and cancer, we measured 15 cell strains containing cytogenetically mapped partial or whole chromosome aneuploidies (Appendix A Table A.1 and Appendix A Figure A.4 and Web Tables E-I). Figure 3.1 shows representative analyses, including detection of whole chromosome gains (Figure 3.1a), a deletion (Fi-



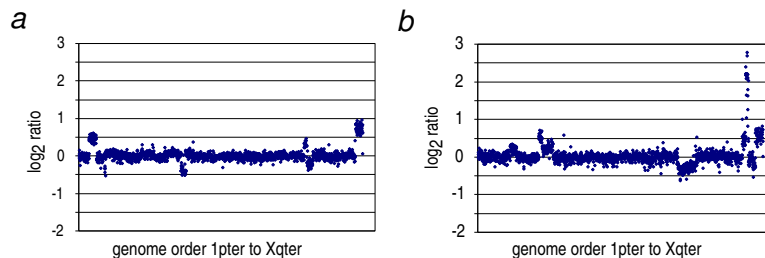
Figure 3.1b) and its confirmation (Figure 3.1c). The mean  $\log_2$ ratios of trisomic chromosomal regions were  $0.49 \pm 0.05$ , compared to the ideal value of 0.58 for a 3/2 ratio. In female/male comparisons, the mean  $\log_2$ ratios on the X chromosome were  $0.72 \pm 0.08$ , compared to the expectation of 1.0. The underestimation of the magnitude of copy number deviations is most likely due to incomplete suppression of repetitive sequences or errors in background subtraction (Pinkel *et al.*, 1998).



**Figure 3.1:** Measurement of single copy changes. **a.** Normalized copy number ratios of a comparison of genomic DNA from cell strain GM03576 and from normal reference DNA (see Web notes for methods). Data are plotted as the mean  $\log_2$ ratio of the triplicate spots for each clone normalized to the genome median  $\log_2$ ratio. The BACs are ordered by position in the genome beginning at 1p and ending at Xq. The borders between chromosomes are indicated with vertical bars. Cytogenetic analysis indicated that this cell line is trisomic for chromosomes 2 and 21. **b.** Normalized copy number variation of cell line GM03563 on BAC clones from chromosome 9. The mean  $\log_2$ ratios of the triplicate spots normalized to the median  $\log_2$ ratio for the genome are plotted relative to the position of the clones on the draft genomic sequence. The  $\log_2$ ratio =  $\sim -1$  indicates a single copy deletion of the first two clones on chromosome 9. The standard deviation of the  $\log_2$ ratios of the clones that are not deleted is 0.08. Colored arrows indicate clones hybridized to interphase nuclei in **c.** **c.** Confirmation of the deletion on 9p by fluorescent in situ hybridization to interphase GM03563 nuclei. The Texas-red labeled test clone RP11-28N06 included in the deletion (indicated by the red arrow in **b**) gave a single red hybridization signal and the FITC labeled reference clone RP11-115L05 (indicated by the green arrow in **b**) gave two green signals.

In principle, each clone might yield a different relationship between ratio and copy number because of the differential ability to block its repetitive sequences. If so, we would expect that ratio differences among clones at the same copy number would become larger as the copy number departed farther from genome average. In the aneuploid cell lines, we found that the vast majority of the autosomal clones had the same response to copy number changes, since the standard deviations of the  $\log_2$ ratios for autosomal clones at 1, 2 or 3 copies all equaled 0.09. However, on the X chromosome, the standard deviation of the  $\log_2$ ratios increased from 0.10 in male/male comparisons to 0.15 in female/male comparisons. Moreover, the ratio variations among X chromosome clones were very reproducible (Appendix A Figure A.3), suggesting that the sequence characteristics of individual clones, possibly differing amounts of sequence shared with the Y chromosome, do have a measurable effect on X chromosome ratios. We detected copy number gains and losses (Figure 3.2a; Web Table J), as well as

amplifications (Figure 3.2b) using DNA isolated from trimmed, frozen breast tumor tissue blocks. Many of the ratio changes are of lesser magnitude than one expects for single copy changes in diploid genomes. For example, the  $\log_2$ ratios of  $0.47 \pm 0.08$  (Figure 3.2a) and  $0.32 \pm 0.07$  (Figure 3.2b) recorded for parts of the genome are less than the expected  $\log_2$ ratio = 0.58 for a copy number ratio of 1.5. These ratios most likely reflect the presence of admixed normal cell DNA, tetraploid DNA content, and/or tumor heterogeneity. In particular, the intermediate ratios indicating a gain of 16p and loss of 16q in Figure 3.2a are likely due to the presence of these aberrations in only a portion of the tumor cells. The magnitude of these easily discriminated ratio changes is well below the "two-fold" level often considered as the limit for significant differences in expression array measurements, indicating the potential of array technology to provide very precise ratio measurements.



**Figure 3.2:** Genome-wide copy number variation in two breast tumors. **a.** and **b.** Normalized fluorescence ratios for breast tumors. We labeled DNA extracted from sections of trimmed frozen tumor specimens with Cy3-dCTP by random priming and hybridized it to the array together with normal male reference DNA (see Web notes for methods). We found low level gains and losses in both tumors and a high level amplification on chromosome 20q in b. The elevated X chromosome ratios reflect the male-female difference in X chromosome copy number. Ratios are plotted as in Figure 3.1a.

Previously, measurement of DNA copy number using arrays assembled from representations of genomic clones prepared by other methods (Geschwind *et al.*, 1998; Lucito *et al.*, 2000) resulted in ratios that were highly variable, requiring averaging over several adjacent clones for detection of single copy changes. In contrast to these other methods, the arrays described here, produced using ligation-mediated PCR representations of the genomic clones, provide reliable data from individual clones, even in polyploid or heterogeneous specimens. Thus, this array CGH platform provides the performance required for potential clinical applications in medical genetic diagnosis and cancer.

### 3.3 Acknowledgements

We thank J. Flint, D. Ledbetter, C. Lese and Vysis, Inc. for telomere clones and clones containing certain named genes. This work was supported by NIH grants CA80314, CA83040, CA84118, HD17665 and CA58207; by California BCRP grant 2RB-0225; and Vysis, Inc.

## Chapter 4

# Mapping segmental and sequence variations among laboratory mice using BAC array CGH

Antoine M. Snijders <sup>1</sup>, Norma Nowak <sup>2</sup>, Bing Huey <sup>1</sup>, Jane Fridlyand <sup>1,4</sup>, Sindy Law <sup>1</sup>, Jeffrey Conroy <sup>2</sup>, Taku Tokuyasu <sup>1</sup>, Kubilay Demir <sup>1</sup>, Readman Chiu <sup>5</sup>, Jian-Hua Mao <sup>5</sup>, Ajay N. Jain <sup>1</sup>, Steven J. M. Jones <sup>5</sup>, Allan Balmain <sup>1</sup>, Daniel Pinkel <sup>1</sup> and Donna G. Albertson <sup>1</sup>

<sup>1</sup>Cancer Research Institute, University of California San Francisco, San Francisco, California 94143

<sup>2</sup>Roswell Park Cancer Institute, Elm and Carlton Streets, Buffalo, New York 14263

<sup>3</sup>Department of Laboratory Medicine, University of California San Francisco, San Francisco, California A 94143

<sup>4</sup>Current address, Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, California 94143

<sup>5</sup>Genome Sequence Centre, British Columbia Cancer Research Centre, 600 West 10th Avenue, Vancouver, British Columbia, Canada V5Z-4E6

## 4.1 Abstract

We used arrays of 2069 BACs (1303 non-redundant autosomal clones) to map sequence variation among *M. spretus*(SPRET/Ei and SPRET/Glasgow) and *M. musculus* (C3H/HeJ, BALB/cJ, 129/J, DBA/2J, NIH, FVB/N and C57BL/6) strains. We identified 80 clones representing 75 autosomal loci of copy number variation ( $(|\log_2\text{ratio}| > 0.4)$ ). These variant loci distinguish laboratory strains. By FISH mapping, we determined that 63 BACs mapped to a single site on C57BL/6J chromosomes, while 17 clones mapped to multiple sites. We also show that small ratio changes ( $\Delta \log_2\text{ratio} \sim 0.1$ ) distinguish homozygous and heterozygous regions of the genome in interspecific backcross mice, providing an efficient method for genotyping progeny of backcrosses.

Supplementary data accompanying this report include array CGH data for mouse strains and species (Web Tables A-J) and backcross mice (Web Table K). Five Supplementary figures are included:

Web Figure A: Ratio of 80 polymorphic BACs in all individuals from all strains.

Web Figure B: FISH mapping results of single site polymorphic BACs.

Web Figure C: FISH mapping result of multi site polymorphic BACs.

Web Figure D: Chromosome distribution of polymorphic BACs.

Web Figure E: Average MAD for all strains and MAD for all individual mice.

## 4.2 Introduction

Natural evolutionary forces and selective inbreeding have given rise to many different mouse strains that exhibit specific characteristics and traits (Beck *et al.*, 2000). The publication of the mouse genome sequence has accelerated investigation of genomic variation among different types of mice. Indeed, genome-wide single-nucleotide polymorphism (SNP) comparison studies across different inbred strains of mice have not only elucidated haplotype structure (Wade *et al.*, 2002; Wiltshire *et al.*, 2003), but have also provided markers for mapping phenotypic differences between strains. Sequence analysis has also identified segmental duplications in genomes, which are defined as stretches of DNA sequence  $> 1\text{-}5$  kb in length with  $> 90\%$  sequence conservation that are present in more than one location in the genome. They make up approximately 5% of the human genome (Bailey *et al.*, 2002) and lesser proportions of rodent genomes, 2.92% in the rat (Tuzun *et al.*, 2004) and 1-1.2% in the mouse (Cheung *et al.*, 2003). These studies have found that both intrachromosomal and interchromosomal or transchromosomal segmental duplications are mosaics of sequences duplicated from within one chromosome or nonhomologous chromosomes, respectively. Transchromosomal duplications are predominantly located in regions close to the centromere and/or telomere and contain few functional genes, whereas intrachromosomal segmental duplications harbor functional genes and gene families and are predominantly found in euchromatic regions (Cheung *et al.*, 2003; Eichler *et al.*, 1996; Horvath *et al.*, 2001; Jackson *et al.*, 1999; Tuzun *et al.*, 2004). Since both types appear to be recent evolutionary events, they are likely to contribute to the generation of phenotypic diversity among laboratory mice. In the course of application of BAC array comparative genomic hybridization (array CGH) to characterize tumors in mouse models, we noticed large strain specific ratio variations for many BACs, indicating probable germline copy number variations. Since these relative increases and decreases in copy number at particular loci could underlie certain strain specific phenotypes,

we undertook a systematic study to detect and map variant loci among laboratory mouse genomes. This approach provides complementary information to genome sequencing, since large-scale copy number differences are not easily identified by sequence analysis (Locke *et al.*, 2003), and it allows investigation of mouse genomes that have not been scheduled for whole genome sequencing. Here, we report on the substantial copy number variation we found among mouse strains. Careful analysis of the results of our inter-strain comparisons indicated that the BAC arrays also provided the ability to detect small ratio variations that presumably reflect nucleotide level variation among the strains. We show that this capability can be used to distinguish chromosomal regions of heterozygosity and homozygosity in interspecific backcross mice. These regions differ by  $\log_2\text{ratio} \sim 0.1$ . Given the high density of genome coverage possible on BAC arrays and the rate at which hybridizations can be performed, this novel application of array CGH provides a highly efficient means to genotype progeny of interspecific backcross mice.

## 4.3 Results

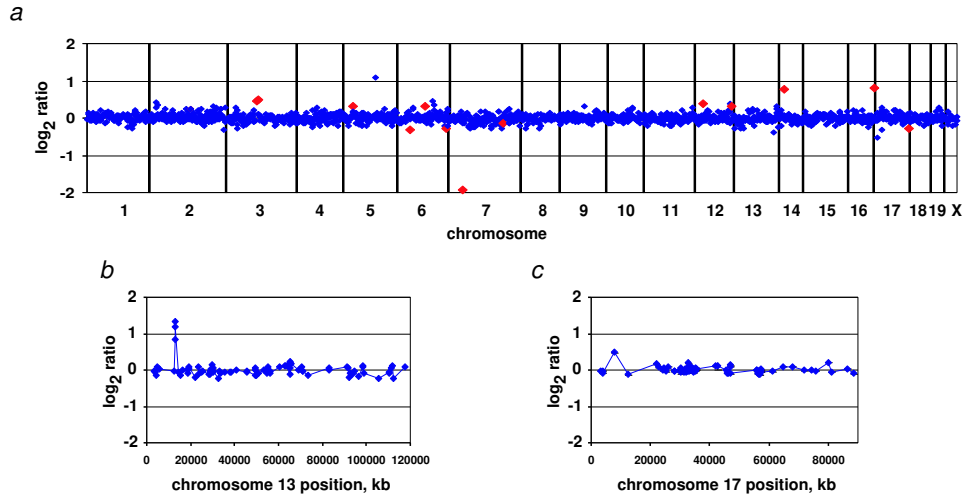
### 4.3.1 Mapping large scale variation in mouse genomes

We used arrays comprised of 2069 mouse BAC clones (1303 unique, non-overlapping autosomal clones) to map loci of segmental DNA copy number variation involving autosomes in multiple different individuals from seven inbred *Mus musculus* strains, including C3H/HeJ (n=5), BALB/cJ (n=4), 129/J (n=4), DBA/2J (n=4), NIH (n=4), FVB/N (n=4), and C57BL/6J (n=9), as well as inbred (SPRET/Ei, n=8) and outbred (SPRET/Glasgow, n=4) *Mus spretus* individuals using genomic DNA from a single FVB/N individual as the reference (WebTables A-J). We identified loci with variant copy number by selecting autosomal clones that gave adequate hybridization signals in at least 95% of all hybridizations and their  $\log_2\text{ratios}$  were  $> 0.4$  or  $< -0.4$  in at least three individuals independent of strain in which they were identified. Eighty clones, corresponding to 75 regions of copy number variation relative to FVB/N, met these criteria (Figure 4.1 and Supplementary Figure A).

We observed that the variant loci represented both strain specific (n = 5) and shared (n = 70) ratio differences across multiple strains or species. Agglomerative hierarchical clustering of all individuals from all strains based on these 75 polymorphic regions (Figure 4.2) was consistent with divergence of the strains and species estimated by other criteria and breeding history (Beck *et al.*, 2000).

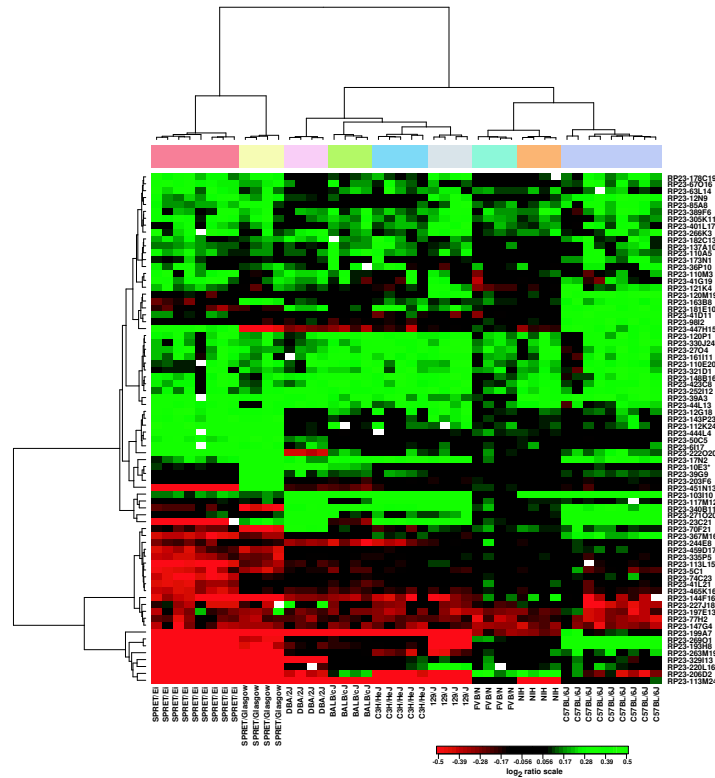
### 4.3.2 Organization of variant loci in the genome

Analysis of duplication content in human, mouse and rat genomes has revealed both a non-random chromosomal distribution and a bias toward clustering of segmental duplications at pericentromeric and subtelomeric regions (Cheung *et al.*, 2003; Eichler *et al.*, 1996; Horvath *et al.*, 2001; Jackson *et al.*, 1999; Thomas *et al.*, 2003; Tuzun *et al.*, 2004). Therefore, to determine whether these murine variant loci identified by array CGH map to particular chromosomal regions, we mapped all 80 polymorphic BACs on C57BL/6J metaphase spreads using FISH. We assessed the distribution of the variant loci by classifying the BAC clones representing these loci as mapping to the proximal 1/3, middle 1/3, or distal 1/3 of the chromosome based on end sequence mapping and FISH position. For this analysis, we considered BACs with overlapping positions in the genome sequence as representative of a



**Figure 4.1:** Array CGH profiles. **a.** Normalized genome-wide DNA copy number profiles of mouse strain DBA/2J using FVB/N genomic DNA as a reference. BACs are ordered by position on the genome starting with chromosome 1 and ending with chromosome X. Vertical bars indicate chromosome boundaries. Each array (MouseArray 3.1) contained 2069 BAC clones, 1848 of which have been mapped onto the draft sequence of the mouse genome (October 2003 freeze). We highlighted in red data points corresponding to a number of the 75 polymorphic clones, which in this individual showed  $|\log_2\text{ratio}| > 0.25$ . **b.** Normalized DNA copy number profiles of chromosomes 13 and 17 after hybridization with SPRET/Glasgow and C57BL/6J genomic DNA, respectively, using FVB/N genomic DNA as a reference. Note the high level DNA copy number polymorphism on the proximal arm of chromosome 13 showing a gain ( $\log_2\text{ratio} \sim 1$ ) spanning three overlapping BAC clones. The proximal part of chromosome 17 shows a lower-level DNA copy number polymorphism ( $\log_2\text{ratio} \sim 0.5$ ) encompassing one BAC clone.

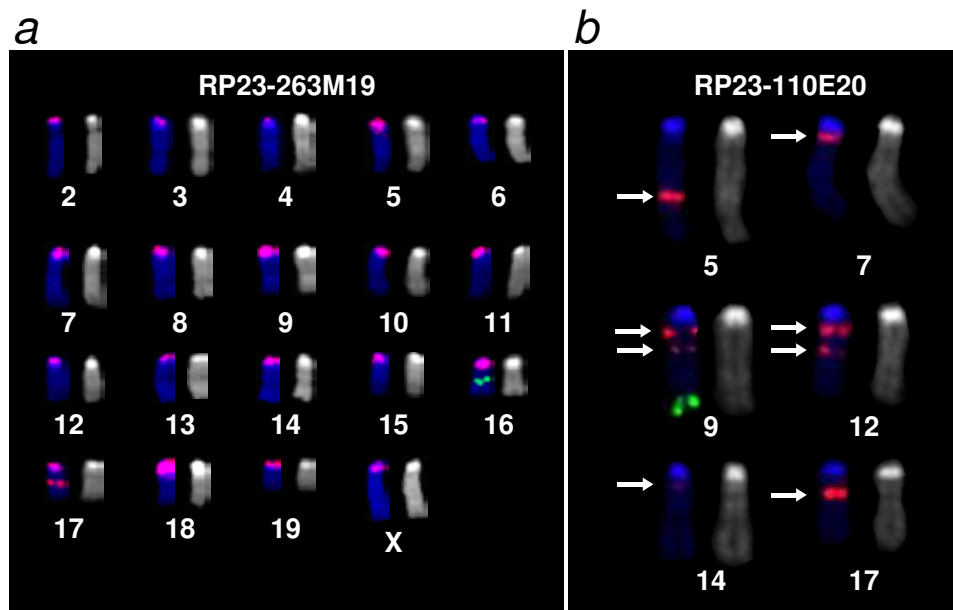
single variant locus. We observed that 63 clones mapped to a single site on C57BL/6J chromosomes (Supplementary Figure B), while 17 clones hybridized to multiple places in the genome (Supplementary Figure C). Two clones, RP23-263M19 and RP23-206D2 mapped to multiple centromeric sites on C57BL/6J chromosomes. Both clones showed a limited number of bands when fingerprinted by *HindIII* digestion, suggesting they might contain large stretches of repeat sequences. RP23-263M19, which was originally assigned to chromosome 16 based on STS content, showed no DNA copy number difference by array CGH when both test and reference genomic DNA originated from *M. musculus*, but a large ratio variation ( $\log_2\text{ratio} < -3$ ) when *M. spretus* genomic DNA was hybridized using *M. musculus* genomic DNA as a reference. This ratio variation is consistent either with deletion of the region in *M. spretus* or amplification of the sequence in *M. musculus*. To distinguish between these possibilities, we mapped RP23-263M19 on metaphase spreads from both *M. musculus* and *M. spretus* (Figure 4.3a and Supplementary Figure C) and found that it hybridized to chromosome 17 in both species, rather than the expected location on chromosome 16 based on the STS associated with this clone. Although the *HindIII* fingerprint for this BAC matched the fingerprint published in the public domain, we could not confirm the presence of the



**Figure 4.2:** Agglomerative hierarchical clustering of 75 regions of sequence variation. We used a Euclidian metric and Ward method to cluster both samples and clones. For overlapping BACs in contigs, we used the one clone from the contig for which the greatest number of observations met the threshold (clone denoted by an asterisk). Note that all mice cluster in two separate branches according to species in agreement with accepted phylogeny. Within the *M. spretus* cluster, all SPRET/Ei mice cluster together and are more closely related to one another than to the outbred SPRET/Glasgow mice. In the *M. musculus* branch, all individual mice from each of the strains cluster together. Furthermore, FVB/N and NIH are closest to each other in the clustering dendrogram, which again agrees with the strain phylogeny.

STS. Thus, we conclude that the clone maps to chromosome 17. In addition to the site on chromosome 17, in *M. musculus* RP23-263M19 also hybridized to all centromeres except 1 and Y (Figure 4.3a). This hybridization pattern does not match the distribution of known repetitive sequences present at the centromeres of mouse chromosomes, including the major and minor satellite DNA (Manuelidis, 1981; Matsuda and Chapman, 1991; Pardue and Gall, 1970; Wong *et al.*, 1990) and a sequence named Amplified Long Genomic Sequence, which maps to all *M. musculus* centromeres except 1, 17 and Y and is not present in *M. spretus* (Koide *et al.*, 1990, 1992). These differences suggest that RP23-263M19 contains a novel centromeric repeat. The second clone that appeared to contain repetitive sequences based on its *HindIII* fingerprint, RP23-206D2 also showed large ratio variation (average  $\log_2$ ratio  $< -0.5$ ) between FVB/N and *M. spretus*, as well as between FVB/N and two other *M. muscu-*

*lus* strains, C3H/HeJ and 129/J. When we hybridized this BAC to C57BL/6J (*M. musculus*) metaphases, it hybridized to the centromeric regions of chromosomes 12, 15, 16, 18 and 19, whereas it hybridized to the telomeric regions of chromosomes 4, 13 and 19 in SPRETUS/Ei (Supplementary Figure C). This hybridization pattern is consistent with the distribution of the rDNA in the genomes of these species and inbred strains (Dev *et al.*, 1977; Eicher and Shown, 1993; Elsevier and Ruddle, 1975; Henderson *et al.*, 1974; Kurihara *et al.*, 1994), suggesting that RP23-206D2 contains some part of the rDNA tandem repeat.



**Figure 4.3:** Cytogenetic mapping of two multi-site BACs. **a.** Overview of C57BL/6J chromosomes showing hybridization signal with BAC clone RP23-263M19 (red) and RP23-140G20 (green), which we used as a marker for chromosome 16. We observed hybridization signals on the middle of chromosome 17 in addition to the centromeres of all chromosomes except 1 and Y, with hybridization signals being brightest on chromosomes 5, 8, 9, 10, 16 and 18 and weakest on chromosomes 3, 4 and 6. **b.** Distribution of hybridization signals on C57BL/6J metaphase spreads from clones RP23-110E20 (red) and RP23-106H14 (green), marker for chromosome 9. RP23-110E20 was selected from the RP23 library by an STS that mapped to 21.6 Mb on chromosome 9. We observed hybridization signals at two sites on chromosome 9 in addition to regions on chromosomes 5, 7, 12 (2 loci), 14 and 17. Note that at least four out of a total of eight hybridization loci showing hybridization with RP23-110E20 are located in the proximal 1/3 of the corresponding chromosomes. The arrows indicate the positions of the RP23-110E20 hybridization signals.

Fifteen other clones that mapped to multiple sites by FISH (excluding the highly repetitive clones RP23-263M19 and RP23-206D2), showed different patterns of ratio variation among the different strains (Figure 4.4). In C57BL/6J, we found these BAC clones hybridized to 21 sites in the proximal 1/3, 15 sites in the middle 1/3 and only seven sites in the distal 1/3 of the chromosomes, with an average number of three signals per BAC (range 2-8). We also

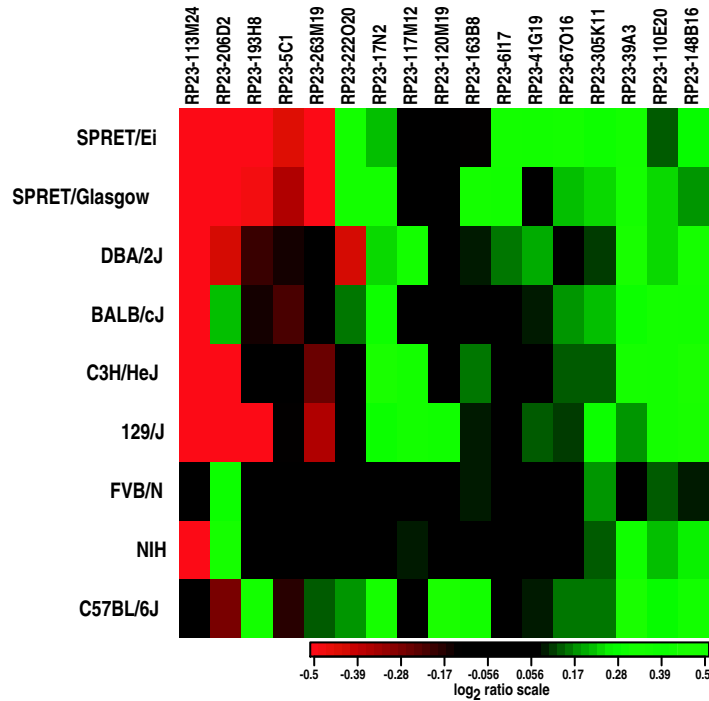


observed that the distribution of hybridization signals was non-random. Moreover, we found that several BACs hybridized to similar sites on the same chromosomes. For example, we observed hybridization more frequently to chromosomes 1 (distal), 5 (middle), 7 (proximal), 9 (proximal and middle), 12 (proximal and middle), 14 (proximal) and 17 (proximal/middle). In some cases, BACs hybridized to more than one of these sites (Figure 4.3b). The fact that most of these BACs hybridize near pericentromeric regions in addition to one or more more distal sites on the chromosome is similar to the reported distribution of segmental duplications in other mammalian genomes (Cheung *et al.*, 2003; Eichler *et al.*, 1996; Horvath *et al.*, 2001; Jackson *et al.*, 1999; Thomas *et al.*, 2003; Tuzun *et al.*, 2004). We found that the set of 53 single site BACs that were correctly mapped with respect to the October 2003 freeze identified loci on all chromosomes except 9, 10 and 16. However, we found that there was not enough evidence to reject the null hypothesis of a uniform distribution of these clones across all chromosomes (Supplementary Figure D). We observed no bias in the position on the chromosome for these clones. Most often, we observed only one hybridization signal by FISH, even in those cases in which the clone showed a ratio difference between C57BL/6J and FVB/N, indicating that it is likely that the ratio deviation between the strains reflects intrachromosomal differences in the number of copies of a sequence at the locus (i.e. too close together to be resolved by FISH) or gain or loss of sequence corresponding to a portion of the BAC. On the other hand, in some cases when we mapped BACs on chromosomes from different strains or species, we observed different numbers of FISH signals, consistent with the ratio variations measured by array CGH (data not shown).

### 4.3.3 Small-scale sequence variation between mouse species allows mapping of genomic content in interspecific backcrosses

In addition to the clones with highly variant ratios representing segmental copy number variation, we observed small ratio variations among all of the clones on the array. Since the magnitude of the variation appeared to depend on which strain was being compared to the FVB/N reference, it seemed unlikely that the variation was due entirely to random fluctuations. Therefore, to quantify the difference among strains, we removed clones with absolute  $\log_2\text{ratio} > 0.3$  to eliminate large ratio variants from the analysis and calculated the average median absolute deviation (MAD) of all clones for all strains (Supplementary Figure E). We observed a range in MAD from 0.07 for FVB/N and NIH strains to 0.12 for SPRET/Ei. We interpret the observed low MAD for the FVB/N and NIH strains as due to the fact that NIH is closely related to FVB/N, which was used as the reference for hybridization. The increase in MAD in the SPRET/Ei, and to a lesser extent in the other strains, is likely attributable to differences in sequence composition, ranging from DNA copy number variations comprising a BAC clone down to single nucleotide differences that affect hybridization efficiency.

Given this observation, we reasoned that BAC array CGH would provide the capability to distinguish strain-specific components of complex genomes, such as those present in interspecific backcross progeny. Therefore, we obtained genomic DNA from three previously characterized mice, which were the progeny of the cross (NIH SPRET/Glasgow) $F_1$  X NIH as well as the (NIH X SPRET/Glasgow) $F_1$  parent. Due to recombination in meiosis, the chromosomes of the backcross mice will have extended regions that are either homozygous NIH or heterozygous NIH/SPRET. The ratio profiles and a statistical analysis for the hybridizations of three of these backcross mice using NIH genomic DNA as the reference are shown in Figure 4.5a-c. We observed that there appeared to be two ratio states differing by  $\log_2\text{ratio}$

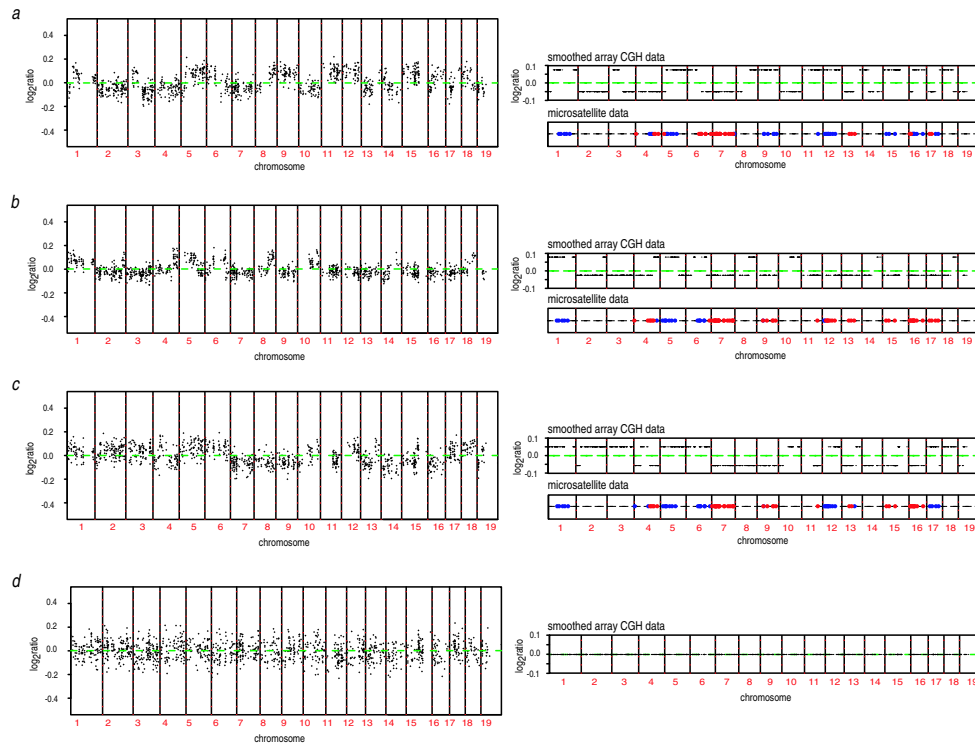


**Figure 4.4:** Copy number differences among mouse strains compared to FVB/N for BACs mapping to multiple sites by FISH. We display the average  $\log_2$  ratio computed over all individuals from each mouse strain or species.

$\sim 0.1$ , with each of the backcross progeny mice having a unique constellation of subtle ratio changes. Hybridization using DNA from the (NIH SPRET/Glasgow) $F_1$  individual with NIH as the reference revealed no ratio changes (Figure 4.5d). We compared the observed ratio changes to available microsatellite data for  $\sim 100$  markers across the genome for all three of the backcross mice and found very high concordance between ratio state and genome composition. Specifically, we found that the higher ratio state corresponded to regions in the genome homozygous for NIH, while regions in the genome showing a lower ratio corresponded to regions heterozygous for NIH (Figure 4.5a-c). Together these observations indicate that array CGH offers a new method for genotyping the progeny of interspecific backcrosses. It allows genome-wide mapping to be accomplished in a single experiment at high resolution and with greater efficiency than conventional methods.

## 4.4 Discussion

Here we have described DNA copy number variants, measured by array CGH, that distinguish the genomes of closely related mouse strains. We observed that 6% of the BAC clones (75 BACs representing polymorphic regions/1303 non-redundant BACs on the array) identified



**Figure 4.5:** Array CGH mapping of interspecific backcrosses. **a., b., c.** Genome-wide DNA copy number profiles of three different individuals from the cross (NIH x SPRET/Glasgow)F1 x NIH using NIH genomic DNA as a reference. We normalized these ratios relative to the ratios from the hybridization of the (NIH x SPRET/Glasgow)F1 parent versus NIH as a reference, in order to minimize fluctuation in ratios due to technical sources. Note the small deviations in copy number ratio around zero (average  $\log_2$ ratio  $< 0.2$ ). The panels on the right show the smoothed genome-wide DNA copy number profiles and microsatellite mapping data for 100 markers across the genome for each of the different mice. Blue indicates regions determined to be homozygous and red regions heterozygous for NIH by microsatellite mapping. We confirmed these observations by carrying out two more hybridizations. First, we repeated the hybridizations for all three mice and found all previously observed ratio changes, and second, we did a dye-swap experiment, which resulted in the expected inversion of the ratio changes (data not shown). Microsatellite mapping agrees with the observed subtle changes in  $\log_2$ ratio and indicates that relative increases in  $\log_2$ ratio are regions homozygous for NIH, while regions of the genome that exhibit a relative decrease in  $\log_2$ ratio correspond to regions heterozygous for NIH. **d.** Genome-wide DNA copy number profile (left) and smoothed genome-wide DNA copy number profile (right) of the (NIH x SPRET/Glasgow)F1 parent using NIH as a reference. Note the absence of subtle copy number deviations as observed in Figure 4.5a-c.

copy number variations in *M. musculus* and *M. spretus* strains relative to FVB/N, with the highest frequency of loci identifying probable intrachromosomal copy number variants (i.e. those identified by single site BACs). However, we expect that these variant sequences make

up much less than 6% of the genome sequence, because the magnitude of the ratio changes, which ranged from our defined threshold of  $|\log_2 \text{ratio}| = 0.4$  to  $|\log_2 \text{ratio}| > 3$ , indicates that only a portion of a BAC is involved in most of the variants. (For example, if the entire BAC were included in a duplicated or deleted region, then one would expect a ratio of 0.8 - 1.0 for the duplication and  $< -3$  for the deletion, since the regions would be homozygous in one strain compared to another.) We also note that ratios on other BACs did not exceed the thresholds, indicative of even smaller segmental duplications, although there is an increased possibility of spurious observations when small ratio changes are measured.

We also found that only 17 of the BACs representing the 75 polymorphic regions hybridized to multiple places in the genome by FISH, which may be due to the nature of the copy number variant loci detectable by array CGH or the composition of the array (Locke *et al.*, 2004). It is likely that regions of high repeat sequence density have been under-represented on this array and such regions may be sources of sequence variation between mouse strains and species, which arguably might be more likely to be distributed at multiple sites in the genome. It is also possible that BACs that map to multiple sites by FISH may be chimeric, containing genomic fragments from different parts of the genome. However, while a low level of chimerism is present in the RP23 library (estimated to be  $\sim 1\%$ ) (Osoegawa *et al.*, 2000), we believe that ratio variation among strains is not due to clone chimerism for a number of reasons. First, chimerism would not affect ratios. Second, we would not have observed that strains and species clustered if the variant loci were randomly formed chimeric clones (Figure 4.2). Third, BAC end-sequencing confirmed the integrity of many of the clones that detected strain-specific variation.

Further research is needed to characterize the variant loci among the strains. Some of the BACs that map to single sites by FISH are likely to identify regions that contain gene families and/or low copy number repeats. For example, *Akr1b3*, *Akr1b8* and *Akr1b7* map within the region spanned by the contig of three polymorphic BACs on chromosome 6. Similarly, BAC end sequencing identifies two positions in the chromosome 17 genome sequence for one of the ends of RP23-11M24 (UCSC Mouse Genome Oct 2003, Supplementary Figure C), indicative of intrachromosomal duplication. Such low copy repeated sequences, may have rendered the region more prone to unequal crossover events and so generated the copy number difference between strains. In the human genome, rearrangements involving segmental duplications are associated with a number of diseases or developmental anomaly syndromes (Emanuel and Shaikh, 2001; Ji *et al.*, 2000; Shaw and Lupski, 2004). Similarly, unequal crossing over between repeated sequences appears to play a major role in the evolution of yeasts grown under selective pressure in the laboratory (Dunham *et al.*, 2002) and in adaptive evolution of sherry wine (flor) yeasts in which copy number variations may comprise as much as 38% of the genome (Infante *et al.*, 2003).

Sequence variations found within an inbred mouse strain include those that arise *de novo*, such as the highly polymorphic pseudoautosomal region (Kipling *et al.*, 1996) and the hypervariable mouse minisatellite locus, Ms6-hm (Kelly *et al.*, 1991). Polymorphisms also are maintained in some inbred strains in spite of many generations of inbreeding (Yuan *et al.*, 1996). We looked for ratio variations within strains for the clones that we initially identified by their variation among strains. However, although we observed some clones showing variation within a strain, we could not rule out technical artifacts. Thus, we conclude that the large-scale variations we detected by array CGH are generally stable in a particular strain.

We also introduced a new application for array based CGH by demonstrating the capability to differentiate regions of heterozygosity and homozygosity in interspecific backcross (NIH

X SPRET/Glasgow) mice. *M. spretus* diverged from *M. musculus* three million years ago, but the two species still remain capable of interbreeding. It has been possible to map a large number of genetic traits in backcrosses using these parental strains (Ewart-Toland *et al.*, 2003; Guenet and Bonhomme, 2003; Nagase *et al.*, 1995; Staelens *et al.*, 2002), because highly divergent strain combinations such as these offer the greatest opportunity to detect phenotypic differences and so maximize success of identifying genes that control specific phenotypes. We note that the approaches described in this report should be applicable to other crosses involving distantly related strains such as *Castaneus* or *Molossinus*, and possibly even to crosses between *Mus musculus* strains.

In the backcross experiments, we observed that regions heterozygous for NIH showed a decrease in  $\log_2$ ratio of about 0.1 compared to homozygous NIH regions in hybridizations using NIH as the reference genome. We propose that the observed higher ratio in regions homozygous for NIH is due to greater sequence conservation between NIH genomic DNA and the C57BL/6J BAC DNA in the array spots compared to the SPRET/Glasgow genomic DNA. On a linear scale, the relative ratio of the heterozygous to homozygous portions of the genome is  $\sim 0.93$ , implying that the SPRET/Glasgow genomic DNA binds to the array elements with about 85-90% the efficiency of NIH. While detection of single nucleotide differences by array hybridization is routine when using short oligonucleotides for array elements, it might at first seem surprising that one can detect such sequence variations on BAC arrays. However, we suggest that it is possible because of a non-uniform distribution of the sequence differences in the genome, resulting in significant reduction in hybridization to a small proportion of the DNA fragments of a BAC. For example, sequencing data (Wade *et al.*, 2002; Wiltshire *et al.*, 2003) and estimates of the degree of sequence divergence compatible with breeding (Sidman and Shaffer, 1994) indicate that the average sequence divergence between the two genomes is on the order of somewhat less than 1%. If we assume that the average length of base pairing for binding to a DNA fragment in an array spot is about 100 bases, then for an average sequence divergence of 1%, one expects approximately 25% of the hybridization sites will have  $> 2\%$  divergence and 8% of the fragments will have  $> 3\%$  divergence if the differences are randomly distributed. It is also possible that the sequence variation may be even more unequally distributed for functional reasons. The fact that the ratios are always lower in the regions heterozygous for NIH and *M. spretus* further supports the view that the subtle ratio variations are due to sequence differences distributed over the genome, rather than resulting from copy number changes of large genomic segments, which, as we have shown, result in both ratio increases and decreases in comparisons between two species. However, it is very likely that there is a continuum of sequence differences between these two types of genomic variation, each with its characteristic effect on comparative hybridizations. Thus, we expect that some of what appears to be "noise" in these measurements will eventually be understood in terms of more subtle genomic characteristics.

## 4.5 Materials and Methods

### 4.5.1 DNA Extraction

We either obtained genomic DNA from different inbred strains of mice from the Jackson Laboratories or isolated genomic DNA from spleens dissected from inbred mice. To isolate DNA, we first froze the spleens on dry ice, and then thinly sliced them. We placed the tissue into a buffer containing 10 mM Tris, 1 mM EDTA, 100 mM NaCl, 1 mg/ml proteinase K

and incubated it overnight at 56°C with shaking. Next, we added RNase to a concentration of 0.25 mg/ml, incubated the mixture at 37°C for one hour, and then incubated the sample again in proteinase K (1 mg/ml) at 56°C for 2 hours. We extracted the preparation with an equal volume of buffer-saturated phenol, collected the aqueous phase and extracted with an equal volume of chloroform-isoamyl alcohol (24:1). We recovered the final aqueous phase; added sodium acetate to a final concentration of 0.3 M followed by two volumes of ice-cold ethanol and inverted the mixture gently until the DNA precipitate appeared. We collected the precipitate by spooling the DNA, washed it in 70% ethanol, and resuspended it in TE. We measured the DNA concentration using fluorometry

### 4.5.2 Preparation of Mouse BAC Arrays

We selected BACs covering the genome from the RP23 C57BL/6J library by two routes; (a) by screening libraries with overgo probes for mapped STS markers as described previously (Cheung *et al.*, 2001) and (b) by identifying BACs with particular STS markers using published information (Hodgson *et al.*, 2001; Cai *et al.*, 1998) or information from the mouse genome sequencing program as the sequencing and radiation hybrid mapping projects progressed. Given the available resources at the time we assembled the array, this procedure resulted in a non-random distribution of BACs among the different chromosomes. In the assembly of the human BAC arrays we reported previously (Snijders *et al.*, 2001), we attempted to minimize the number of potentially polymorphic or multi-site clones on the array by screening using both FISH (Cheung *et al.*, 2001) and array CGH with known aneuploid cell lines (Snijders *et al.*, 2001). However, for the mouse array, we included all selected BACs on the array and subsequently carried out FISH mapping to identify multi-site clones or potentially erroneously mapped clones. Thus, while there is some bias in the representation of the chromosomes on the arrays, we have not intentionally selected against BAC clones representing regions of sequence variation among strains. Information on the clone set is provided at: <http://microarrays.roswellpark.org>. We prepared DNA spotting solutions from the BACs by ligation-mediated PCR as described previously for assembly of human BAC arrays (Snijders *et al.*, 2001). We printed each BAC clone in triplicate, with spots on 130  $\mu\text{m}$  centers in 12 x 12 mm area on chromium surfaces using a custom robot.

### 4.5.3 DNA labeling by random priming

For each labeling reaction, we used approximately 300 ng each of test or reference genomic DNA in a volume of 15  $\mu\text{l}$  containing 10 mM Tris-1 mM EDTA (pH7.5), 0.2 mM unlabeled dATP, dCTP, and dGTP, 0.1 mM unlabeled dTTP, 1x random primer (Bioprime DNA labeling system, Invitrogen) and 0.4 mM Cyanine-3 (test) or Cyanine-5 (reference) conjugated dUTP (Amersham). We denatured the DNA in a thermal cycler at 100°C for 15 min. and cooled it to 4°C before adding 12 units of Klenow fragment (Bioprime DNA labeling system, Invitrogen). Using a thermal cycler we incubated the reaction at 16°C for 10 min. and 37°C for 20 min. for 30-40 cycles. We removed unincorporated nucleotides from the labeling reaction using a Sephadex G-50 spin column and visually assessed the labeling efficiency by the intensity of the color of the flow-through.

#### 4.5.4 Array hybridization

For each hybridization, we combined Cy3-dUTP labeled test DNA and Cy5-dUTP labeled reference DNA with 40  $\mu$ g mouse Cot-1 DNA (Invitrogen) and precipitated the mixture using sodium acetate at 0.3 M and two volumes of ice-cold ethanol. We chilled the mixture at  $-20^{\circ}\text{C}$  for 20 min. and collected the precipitate by centrifugation at  $17400 \times g$  for 25 min. We removed the supernatant, wicked dry the pellet with a tissue and resuspended the pellet in a 40  $\mu$ l hybridization solution consisting of 50% formamide, 10% dextran sulfate (mw  $\sim$ 500,000), 2X SSC, 4% SDS and water. To denature the probe mixture, we heated it at  $75^{\circ}\text{C}$  for 20 min and then incubated it for several hours at  $37^{\circ}\text{C}$  to allow pre-annealing of the Cot-1 DNA to the probe DNA. We applied a ring of rubber cement around the perimeter of each array to contain the hybridization mixture. We added 50  $\mu$ l hybridization solution containing no probe DNA within the perimeter of the rubber cement to pre-wet the array for 15 min. We aspirated the wetting solution and added the hybridization mixture to the array and then placed the slide into a polyethylene slide-mailing container containing 40  $\mu$ l of 50% formamide, 2X SSC to maintain humidity, sealed the container with Parafilm and incubated it with rocking at  $37^{\circ}\text{C}$  for 60 hours. After hybridization, we rinsed off excess hybridization mixture using PN buffer (0.1 M  $\text{Na}_2\text{HPO}_4/\text{NaH}_2\text{PO}_4$ , pH 8.5, 0.1% NP-40) and washed the arrays at  $45^{\circ}\text{C}$  in the following order: 50% formamide, 2X SSC for 20 min., 2X SSC for 10 min., 0.1X SSC for 10 min., PN buffer for 10 min., and 0.1X SSC for 10 min. We removed the rubber cement borders and drained off excess liquid. We wet-mounted the arrays under a cover slip in a solution containing 90% glycerol, 10% PBS and 1  $\mu$ M DAPI. We stabilized the cover slip using a border of fingernail polish along two edges.

#### 4.5.5 Imaging and Analysis

We acquired 16 bit 1024x1024 pixel DAPI, Cy3 and Cy5 images using a custom built CCD camera system as described previously (Pinkel *et al.*, 1998) and used "UCSF SPOT" software (Jain *et al.*, 2001) to automatically segment the spots based on the DAPI images, perform local background correction and to calculate various measurement parameters, including  $\log_2$ ratios of the total integrated Cy3 and Cy5 intensities for each spot. We used a second custom program SPROC to obtain averaged ratios of the triplicate spots for each clone, standard deviations of the triplicates and plotting position for the BACs on the October 2003 freeze of the mouse genome sequence (<http://genome.ucsc.edu>). We edited the data files to remove ratios on clones for which only one of the triplicates remained after SPROC analysis and/or the standard deviation of the  $\log_2$ ratios of the triplicates was  $> 0.2$ . We averaged the values for replicate clones and removed exact replicates from the dataset. We created two different versions of the dataset. For the copy number polymorphism analysis, we included only autosomal clones that were present in at least 95% of the hybridizations. We included clones that were present in at least 75% of the samples when plotting ratios along the genome and in the analysis of heterozygosity and homozygosity in backcross animals.

#### 4.5.6 Statistical methods

We identified BAC clones representing putative copy number variants using a conservative threshold of  $|\log_2\text{ratio}| = 0.4$  in at least three samples. When contiguous clones were identified, we retained the one clone from the contig for which the greatest number of observations

met the threshold. We used the resulting set of clones to cluster the samples using an agglomerative hierarchical clustering algorithm with Euclidean metric and Ward method. We re-ordered the samples within each branch according to their means in such a way that tree topology was preserved.

We estimated the variability of the individual hybridizations by computing the Median Absolute Deviation (MAD) of all clones for a given sample with  $|\log_2 \text{ratio}| < 0.3$ . We considered the clones that had absolute mean value above 0.3 in a given strain and less than 0.1 in the remaining strains to represent possible copy number polymorphisms specific to that strain, even though they might not meet the more conservative criteria described above. We tested the uniformity of the distribution of the mapped clones and of the single-site polymorphism clones among chromosomes using the chi-square goodness of fit test (Snedecor and Cochran, 1989). Specifically, we tested the equality of the number of the mapped clones per Mb on different chromosomes, as well as the equality of the proportion of the single-site polymorphic clones among clones mapped on the genome sequence of different chromosomes.

To identify heterozygous and homozygous regions in the genomes of backcross animals, we first normalized each hybridization to the F1 hybridization in order to minimize systematic ratio fluctuations due to technical sources. We then applied a modification of the unsupervised Hidden Markov Model (HMM) state-fitting method described previously 41. Here, we restricted the maximum number of states to be two and fit an HMM to all chromosomes simultaneously by concatenating the neighboring chromosomes and assigning values to either state 1 or 2. We selected the number of states (1 or 2) using a BIC criterion with  $\delta = 1.5$  42-44. We merged the states if their median values were less than 0.05 apart. For each hybridization, we plotted the smoothed value of all clones that were not aberrations or outliers. The smoothed value is computed as the median value of the corresponding state for a given clone. In the case of one state only, we used the median value of all clones for a given hybridization. We note that the unsupervised HMM procedure used to assign individual BACs to a given state (Figure 4.5) does so by assigning a probability to each BAC of being in a given state and then allocates each BAC to the state with the largest probability (Fridlyand *et al.*, 2004). Thus, a natural output of the procedure is a measure of the confidence with which a BAC is assigned as heterozygous or homozygous, which allowed us to determine that the majority of BACs were assigned to their respective states with >95% confidence.

#### 4.5.7 BAC end sequencing analysis

We subjected the end sequences to BLAT analysis on the October 2003 freeze of the mouse genome (UCSC Genome Browser). We considered a BAC to be properly aligned on the genome sequence if the sequence pairs from both end-sequences from any one BAC were on the same chromosome, no more than 400 kb apart on opposite strands. We determined that 64 of the clones mapped correctly based on concordance of our FISH and/or BAC end sequence data with position in the October 2003 freeze as assigned by BAC end pairs or STS position. However, we found that the positions of 13 clones were discrepant and another three were not mapped on the October 2003 freeze. We found no evidence of chimerism based on the positions of the BAC end sequences on the October 2003 freeze.



### 4.5.8 Fluorescent *in situ* hybridization (FISH)

In order to verify the identity of the variant loci, we mapped all 80 polymorphic BACs on C57BL/6J metaphase spreads using FISH. We carried out two-color FISH using two different BAC DNAs, one labeled with Cy3-dUTP and the other with fluorescein-dCTP, so that a control probe hybridizing to a known chromosomal region could be incorporated into the same hybridization as the test probe. We labeled 20 ng of BAC DNA in a 10  $\mu$ l reaction containing 0.2 mM each of unlabeled dATP, dCTP, and dGTP and 0.1 mM dTTP if cy3-dUTP was the label, or 0.2 mM unlabeled dATP, dGTP, and dTTP and 0.1 mM dCTP if fluorescein-dCTP was the label, 1x random primer (Bioprime DNA labeling system, Invitrogen) and 0.4 mM Cy3-dUTP or fluorescein-dCTP. We denatured the BAC DNA along with random primers and nucleotides at 100°C for 15 min and then cooled the mixture to 4°C after which we added 20 units of Klenow fragment (Bioprime DNA labeling kit, Invitrogen) and incubated the reaction at 37°C for ~16 hours. We combined 5  $\mu$ l each of a Cy3 and a fluorescein labeling reaction with 6  $\mu$ g of mouse Cot-1 DNA and precipitated the mixture using sodium acetate at 0.3 M and two volumes of ice-cold ethanol. We removed the supernatant and wicked dry the pellet with a tissue and then resuspended it in 10  $\mu$ l of a hybridization solution consisting of 50% formamide, 10% dextran sulfate (mw ~500,000), 2X SSC, and 2% SDS. We then denatured the hybridization solution at 75°C for 20 min. To prepare metaphases for hybridization, we washed microscope slides containing metaphase spreads from *M. musculus* or *M. spretus* embryonic fibroblasts in 2X SSC at 37°C for 30 min., dehydrated them in an ethanol series and allowed them to air dry. We then denatured the metaphase spreads in 70% formamide, 2X SSC (pH 7.0) at 75°C for 4 min, dehydrated them again in an ethanol series, and allowed them to air dry. After warming the metaphase spreads to 37°C we applied the denatured probe to the metaphases. We hybridized the probes under a cover slip in a humidified chamber at 37°C for 40 hours. We then removed the cover slip and excess hybridization solution by immersing the slide in PN buffer and washed the arrays at 45°C in the following order: 50% formamide, 2X SSC for 20 min., 2X SSC for 10 min., 0.1X SSC for 10 min., PN buffer for 10 min., and 0.1X SSC for 10 min. After removing excess liquid we cover slip mounted the slide in Vectashield (Vector Laboratories) and used a CCD based imaging system to capture three color images.

## 4.6 Acknowledgements

We thank members of the Albertson and Pinkel laboratories, who participated in the assembly and printing of the mouse BAC arrays, Gillian Hirst for providing mouse embryo fibroblasts used for preparation of metaphase spreads, Lawrence Hon for help with BAC end sequence analysis and Karen Kimura and the UCSF Cancer Center Informatics Core for maintenance of the microarray database. We also thank Susan Deveau of the Jackson Laboratory for her assistance in obtaining DNA from mouse strains. Gregg Magrane and the UCSF Immunohistochemistry and Molecular Pathology Core assisted with FISH mapping of some of the array clones. This work was supported by NIH/NCI grants CA84118 and P30 CA16056 and a Supplement P02137165CN from the NCI Mouse Models of Human Cancer Consortium.



## Chapter 5

# Shaping of Tumor and Drug Resistant Genomes by Instability and Selection

Antoine M. Snijders<sup>1,3</sup>, Jane Fridlyand<sup>1,3</sup>, Dorus A. Mans<sup>1,3</sup>, Richard Seagraves<sup>2,3</sup>, Ajay N. Jain<sup>1,2,3</sup>, Daniel Pinkel<sup>2,3</sup> and Donna G. Albertson<sup>1,2,3</sup>

<sup>1</sup>Cancer Research Institute, University of California San Francisco, San Francisco CA, USA

<sup>2</sup>Department of Laboratory Medicine, University of California San Francisco, San Francisco CA, USA

<sup>3</sup>University of California San Francisco Comprehensive Cancer Center, San Francisco, San Francisco CA, USA

## 5.1 Abstract

Tumors with defects in mismatch repair (MMR) show fewer chromosomal changes by cytogenetic analyses than most solid tumors, suggesting that a greater proportion of the alterations required for malignancy occur in genes with nucleotide sequences susceptible to errors normally corrected by MMR. Here, we used genome-wide microarray comparative genomic hybridization to carry out a higher resolution evaluation of the effect of MMR competence on genomic alterations occurring in 20 cell lines and to determine if characteristic aberrations arise in MMR proficient and deficient HCT116 cells undergoing selection for methotrexate resistance. We observed different spectra of aberrations in MMR proficient compared to deficient cell lines, as well as among cell lines with different types of MMR deficiency. We also observed different genetic routes to drug resistance. Resistant MMR deficient cells most frequently displayed no copy number alterations (16/29 cell pools), whereas all MMR proficient cells had unique abnormalities involving chromosome 5, including amplicons centered on the target gene, *DHFR* and/or a neighboring novel locus (7/13 pools). These observations support the concept that tumor genomes are shaped by selection for alterations that promote survival and growth advantage, as well as by the particular dysfunctions in genes responsible for maintenance of genetic integrity.

## 5.2 Introduction

The development of solid tumors is associated with the acquisition of genetic and epigenetic alterations and corresponding changes in gene expression that modify normal growth control and survival pathways. These changes may be brought about at the genomic level in a variety of ways, including, for example, altered karyotypes, point mutations and epigenetic mechanisms. It is now generally agreed that in order for a sufficient number of alterations to accumulate to cause a malignancy, one or more of the mechanisms that work to maintain genetic integrity in cells and/or to regulate cell cycle progression must be compromised, presumably through mutations that occur early in tumorigenesis (Loeb, 2001). Genomic DNA copy number aberrations are frequent in solid tumors and are expected to contribute to tumor evolution by copy number induced alterations in gene expression. A variety of cytogenetic and more recently array based (Snijders *et al.*, 2001) analytic methods have found a wide range in the number and types of chromosome level alterations present in human tumors, which are likely to reflect the many different solutions taken by individual tumors to escape normal protective mechanisms. These analyses also show that the genomes of established tumors are remarkably stable, as evidenced by the similarity of tumor recurrences with primaries (Waldman *et al.*, 2000; Albertson, 2003), indicating that the set of aberrations is maintained due to continued selective advantage. Taken together these observations suggest that the spectrum of alterations that one sees in fully developed tumors is a composite of selection acting on the variation that is permitted to arise by the particular failures in genomic surveillance mechanism(s) present in the tumor. This leads, for example, to the expression of the same gene being altered in multiple ways in different tumors, or particular functional pathways being affected at different locations because certain genes in the pathways are more susceptible to alteration by the failures in genomic surveillance present in that tumor.

The interplay between selection and genetic instability in shaping tumor genomes is currently most clearly established in tumors with defects in mismatch repair. These tumor genomes show a high level of microsatellite instability due to failures in mismatch repair (MMR) genes

(generally *MSH2* or *MLH1*) either through mutation, as in tumors from patients with hereditary non-polyposis colorectal cancer (Bocker *et al.*, 1999), or in sporadic tumors through silencing of *MLH1* (Esteller, 2000). The MMR deficient colorectal tumors also differ from MMR proficient tumors in their histology (Bocker *et al.*, 1999), in the genes that are inactivated in the tumors (Zhang *et al.*, 2001a; Duval and Hamelin, 2002), in their response to therapy (Aebi *et al.*, 1996; Fink *et al.*, 1998) and by the relatively low number of chromosomal level alterations in their genomes (Muleris and Dutrillaux, 1990; Remvikos *et al.*, 1995; Schlegel *et al.*, 1995; Eshleman *et al.*, 1998; Soulie *et al.*, 1999). Thus MMR deficient tumors can be readily distinguished from MMR proficient tumors based on the types of genetic instability they display and the resultant selection for alterations in those critical genes that are susceptible to mutation through microsatellite instability.

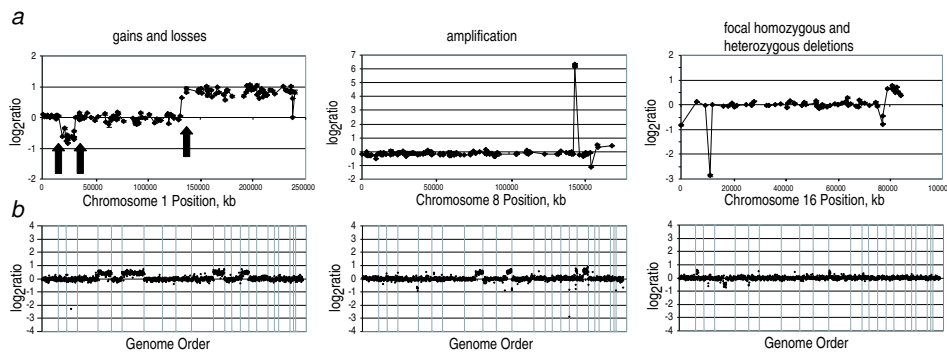
The relation of mechanistic defect to aberration type has not been established for most sporadic tumor types, but it is likely that some of the variety and complexity associated with these tumor genomes might be rationalized if associations could be developed between particular aberrations and specific defects in maintenance of genome stability. The recently developed high throughput array based genome scanning techniques such as microarray-based comparative genomic hybridization (CGH) now offer the means to analyze many tumor genomes and to more precisely define their genomic alterations. The higher resolution and precision afforded by these measurements provide the opportunity to develop approaches to the interpretation of DNA copy number alterations that recognize particular types of genetic instability. Here, we demonstrate this possibility with a study of the genomic effects of MMR competence. Although it has been known for some time that tumors with defects in mismatch repair show few chromosome level changes compared to most sporadic solid tumors as measured by cytogenetics, flow sorting and chromosome CGH, the data we present here not only confirm the earlier cytogenetic analyses, but also reveal that there are differences in aberration frequencies associated with the particular MMR gene defects. In order to investigate the effect of MMR competence on genome evolution in more detail, we extended these analyses to a model system in which we could repeatedly evaluate the genomic changes arising in response to selection for resistance to a single agent in cell populations with defined genetic backgrounds. The analysis of multiple independent methotrexate resistant cell pools revealed differences in aberrations associated with MMR proficient and deficient HCT116 cells. Furthermore, they showed that when copy number changes were present, they included plausible drug resistance target genes, indicating that the frequency of aberrations that could be considered bystanders was low. These studies provide a model system and framework for further aberration-based classification of tumor genomes.

## 5.3 Results

### 5.3.1 Copy number aberrations in MMR deficient and proficient cell lines

Cytogenetic analyses have shown that tumors with defects in mismatch repair (MMR) have fewer chromosomal changes than most solid tumors, suggesting that a greater proportion of the alterations required for malignancy occur in genes whose nucleotide sequences are susceptible to errors normally corrected by this system. Here, we used genome-wide array CGH (Snijders *et al.*, 2001) to carry out a higher resolution evaluation of the effect of MMR competence on the genomic alterations occurring in 10 MMR deficient cell lines (see Ap-

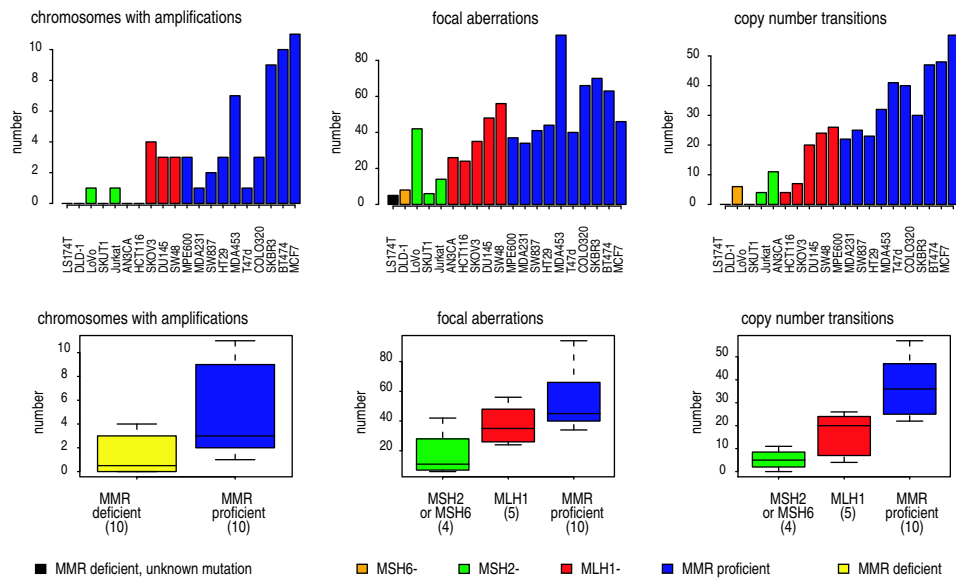
pendix B Figure B.1); complete data sets available at: <http://cc.ucsf.edu/albertson/public/>) compared to 10 MMR proficient cell lines (see Appendix B Figure B.2 and (Snijders *et al.*, 2001)). We counted the number of gains or losses of whole chromosomes, which might be expected to occur following failures of karyokinesis or cytokinesis, and the number of copy number transitions within a chromosome, which are likely to reflect DNA strand breakage that led to non-reciprocal translocations. We also distinguished two types of focal aberrations, those involving low level gains or losses of DNA sequence spanning one to two clones and gene amplifications, which we defined as focal high level copy number changes (Figure 5.1a). We found few copy number alterations in MMR deficient cells (Figure 5.1b). The MMR deficient cells showed significantly fewer aberrations of all types compared to MMR proficient cells (Figure 5.2) in accord with earlier observations, although we observed a substantial number of aberrations in some MMR deficient lines. We also found a dependence of aberration type on the specific MMR defect. Cells deficient in *MLH1* had a higher frequency of copy number transitions and focal aberrations than *MSH2* deficient cells (Figure 5.2).



**Figure 5.1:** Copy number profile phenotypes. **a.** Definition of aberration types. Copy number profiles of individual chromosomes showing from left to right gains and losses of portions of chromosomes or chromosome arms (arrows indicate copy number transitions at the boundaries of the loss and gain), amplification, and focal homozygous ( $\log_2\text{ratio} \sim -3$ ) and heterozygous deletions ( $\log_2\text{ratio} \sim -1$ ). **b.** Array CGH profiles of MMR deficient cell lines. Whole genome scans of MMR deficient cell lines compared to normal male reference DNA. Data are plotted as the mean  $\log_2\text{ratio}$  of the triplicate spots for each clone normalized to the genome median  $\log_2\text{ratio}$ . The clones are ordered by position in the genome beginning at 1p and ending with Xq (if measured on HumArray1.14) or Y (if measured on HumArray2.0). The borders between chromosomes are indicated with vertical bars. Cell lines from left to right are Jurkat (*MSH2*-), HCT116 (*MLH1*-) and DLD-1 (*MSH6*-). Complete data sets are available at: <http://cc.ucsf.edu/albertson/public/>.

### 5.3.2 Copy number aberrations in methotrexate resistant MMR deficient and proficient cells

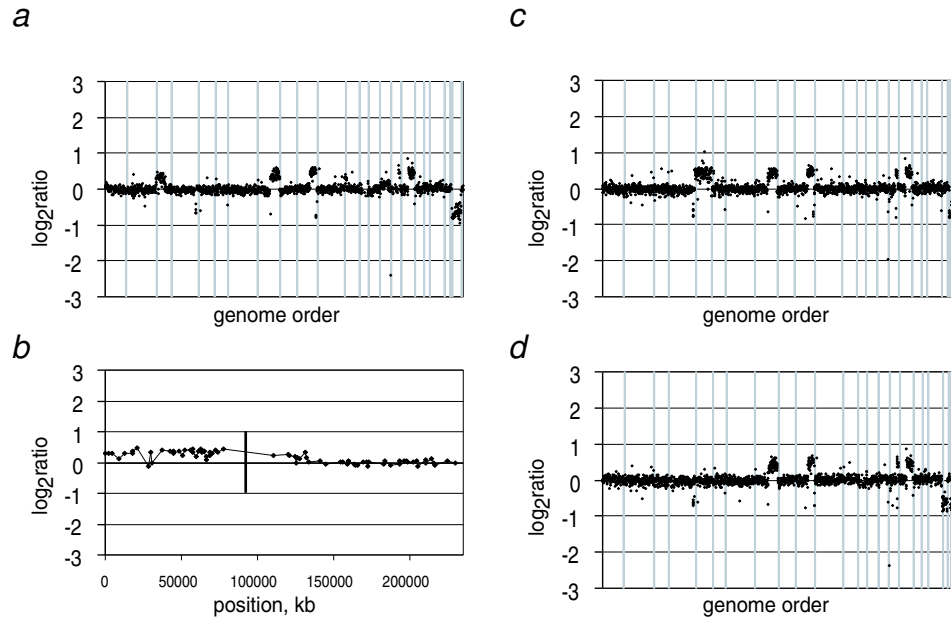
We further investigated the influence of MMR competence on genome evolution, by analyzing the copy number changes that occurred in multiple independent populations of methotrexate resistant cells arising in MMR deficient and proficient cells. We chose methotrexate re-



**Figure 5.2:** Frequencies of aberration types in mismatch repair deficient and proficient cell lines. **a.**

Aberration types. The numbers of copy number transitions, amplifications, focal deletions and whole chromosome gains or losses were recorded for each cell line. Array CGH data were reported previously for mismatch repair proficient cell lines (Snijders *et al.*, 2001). For mismatch repair deficient cell lines reported in this publication, the complete data sets are available at: <http://cc.ucsf.edu/albertson/public/>. For these analyses, we grouped the one *MSH6* deficient cell line with the *MSH2* deficient lines. Mismatch repair proficient cells tended to have a greater number of each aberration type, except whole chromosome changes. Cells deficient in *MLH1* contained more transitions and focal aberrations than *MSH2* (and *MSH6*) deficient cells, and the numbers were similar to some MMR proficient cells. **b.** Distributions of copy number aberration types in MMR proficient and deficient cells. The number of chromosomes containing amplifications was significantly higher in MMR proficient cells compared to MMR deficient cells ( $p < 0.02$ ). Cells proficient for MMR contained significantly more transitions than MMR deficient cells ( $p < 0.02$ ) and more focal aberrations ( $p < 0.02$ ).

sistance as the model system, because clinical resistance to methotrexate targets a number of genes by a variety of mechanisms, including mutations at the nucleotide level and gene amplification (see below), thereby providing the opportunity to determine which types of aberration occur more frequently in MMR deficient and proficient backgrounds. We used MMR deficient HCT116 (*MLH1*-) and HCT116+chr3 cells made MMR proficient by introduction of chromosome 3 carrying a wild type copy of *MLH1* (Koi *et al.*, 1994; Hawn *et al.*, 1995). Array CGH analysis of the HCT116+chr3 cells indicated that they carry an additional copy of most of 3p (Figure 5.3a and b). Fluorescent *in situ* hybridization with probes to 3p and 3q confirmed the presence of two normal copies of chromosome 3 and an additional telocentric chromosome showing only hybridization to the 3p probe (data not shown). Slightly elevated copy numbers of chromosomes 12p and 15 are also evident, indicating additional copies in a minority of the cells (Figure 5.3a).

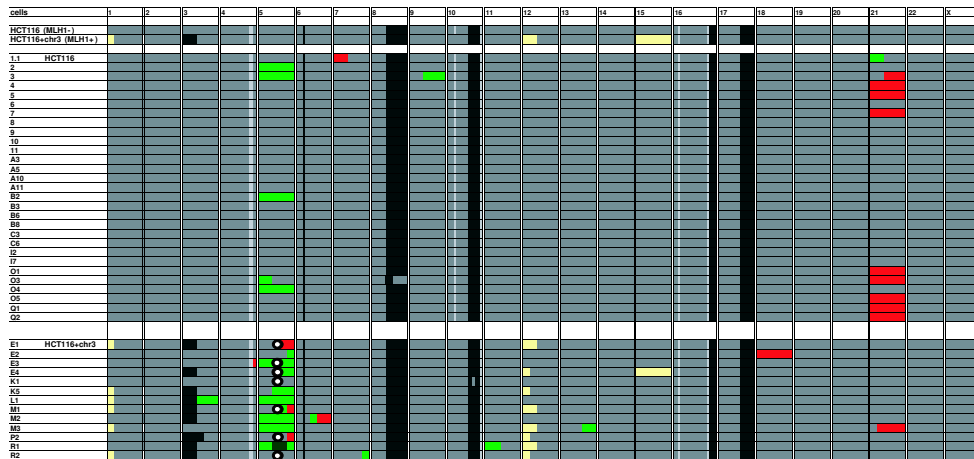


**Figure 5.3:** Copy number profiles of untreated and methotrexate resistant pools of HCT116 and HCT116+chr3 cells compared to normal male reference DNA. **a.** Whole genome scan of HCT116+chr3 cells. The HCT116+chr3 cells show a single copy increase on chromosome 3p in addition to the copy number aberrations present in HCT116. Data are plotted as in Figure 5.1. The complete data set is available at: <http://cc.ucsf.edu/albertson/public/>. **b.** Copy number profile of chromosome 3 in HCT116+chr3 cells. The normalized copy number is plotted in order from 3p to 3q on the August 2001 draft sequence and shows the characteristic discontinuous copy number increase on 3p in these cells. The vertical bar indicates the position of the centromere. **c.** Whole genome scan of a methotrexate resistant pool of HCT116 cells. The copy number profile shows a gain of chromosome 5 in addition to the copy number changes characteristic of HCT116. Data are plotted as in Figure 5.1. **d.** Whole genome scan of a methotrexate resistant HCT116 cell pool. The resistant cells show the loss of one copy of chromosome 21 in addition to the copy number aberrations present in the untreated HCT116 cells. Data are plotted as in Figure 5.1.

Genome-wide analysis of copy number aberrations in methotrexate resistant cell pools (Figure 5.4) indicated that MMR deficient HCT116 cells most frequently showed no copy number aberrations (16/29 pools). If copy number aberrations were present, they involved loss or gain of whole chromosomes, including loss of one copy of chromosome 21 (9/29 pools, Figure 5.3d), or gain of a copy of chromosome 5 (4/29 pools, Figure 5.3c). We rarely observed aberrations involving partial chromosome arms (5 copy number transitions/29 pools). In contrast to a previous report (Lin *et al.*, 2000), we observed no amplifications of *DHFR*, even when the level of methotrexate used for selection was increased from 25 nM to 100 nM (data not shown).

A different spectrum of aberrations occurred in MMR proficient HCT116+chr3 cells. All resistant cell pools showed aberrations involving chromosome 5 (Figure 5.4). Most frequently,

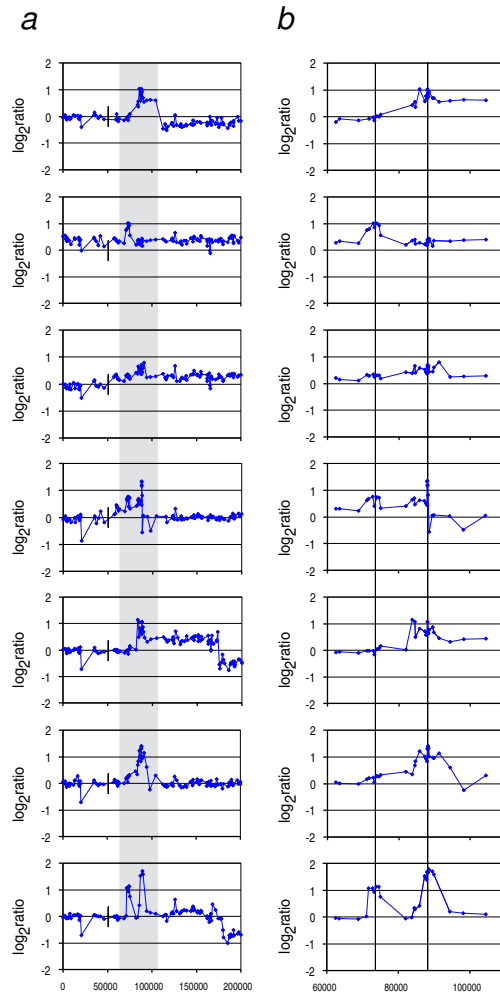




**Figure 5.4:** Schematic overview of copy number aberrations in methotrexate resistant HCT116 (MMR deficient) and HCT116+chr3 (MMR proficient) cell pools as determined by array CGH. Horizontal bars indicate different pools, while vertical bars indicate individual chromosomes. Copy number aberrations characteristic of HCT116 and HCT116+chr3 are indicated in dark gray and light gray for copy number gains and losses, respectively. The untreated HCT116+chr3 cells show variable low level gains of 1p, 12p and 15, indicating that the population is heterogeneous with respect to these aberrations. In the resistant cell pools, the presence and extent of these copy number changes is indicated in yellow. Acquired copy number aberrations in the methotrexate resistant cell pools are indicated in green and red for copy number gains and losses, respectively and with a colored circle for amplifications. The MMR deficient cells when selected with 25 nM methotrexate frequently show no copy number aberrations, loss of an entire chromosome 21 or a gain of an entire chromosome 5. All MMR proficient cells selected with 25 nM methotrexate show unique copy number aberrations involving chromosome 5, including amplification of the *DHFR* locus and/or a genomic region proximal of the *DHFR* locus. One cell pool, R1 showed a higher level gain of the 5q14 region, which is indicated in dark green. Complete data sets are available for resistant cell pools with copy number changes at: <http://cc.ucsf.edu/albertson/public/>.

we observed amplicons (7/13 pools) centered at the *DHFR* locus and/or a neighboring 13 Mb region flanked by RP11-174I22 and RP11-172K14 that mapped proximal to *DHFR* (Figure 5.5). These recurrent amplicons did not include *HMGCR*, previously reported to be co-amplified with *DHFR* in methotrexate resistant cells (Srimatkandada *et al.*, 2000). We also observed that the copy number profile for all of chromosome 5, as well as within the amplified region was unique to each resistant cell pool. For example, in three cases there were losses of portions of 5q distal to the amplicon, but the extent of the deleted region varied. In other cases, a gain of all or a portion of 5q was present in addition to the amplicon(s). In the resistant HCT116+chr3 cell pools without amplification, we observed gains of chromosome 5 or 5q (5/13 pools) together with a loss of distal chromosome 21 in one case (Figure 5.4). We observed a non-recurrent pattern of aberrations in one resistant cell pool, which included a gain of distal 5q and loss of chromosome 18 (Figure 5.4). In contrast to MMR deficient cells, we observed a higher frequency of aberrations involving partial chromosome arms (7 copy number transitions/13 pools, excluding copy number changes on chromosome 5 when

amplifications were present).



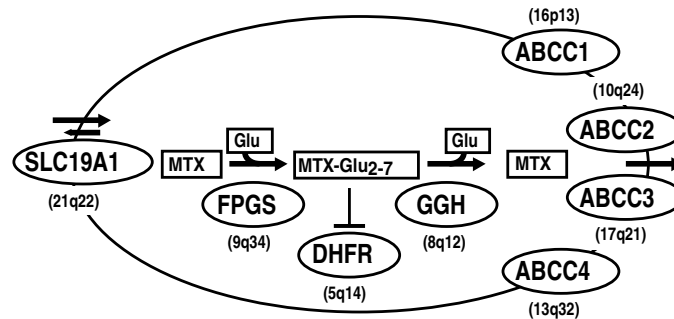
**Figure 5.5:** Copy number aberrations on chromosome 5 in methotrexate resistant HCT116+chr3 cell pools. **a.** Normalized  $\log_2$  copy number ratios of clones on chromosome 5 measured on HumArray2.0 and the higher resolution array spanning the *DHFR* region. The normalized copy number is plotted in order from 5p to 5q according to position on the August 2001 draft genome sequence. All cell pools show low level copy number aberrations of parts of chromosome 5, as well as regions of amplification at 5q14. The gray bar indicates the region of recurrent amplification. **b.** Expanded view of the region of recurrent amplification. Vertical lines are drawn through two regions of recurrent amplification and pass through the *DHFR* locus and a proximal locus. In all cell pools, *DHFR* is included in either a low-level gain or is amplified. Moreover, in several cell pools a region proximal to *DHFR* is amplified together with a low level gain or an amplification of *DHFR*. No genes previously associated with methotrexate resistance map to the proximal region.

Since studies of the organization of amplicons by cytogenetics indicate that the amplified DNA may be present in multiple copies on a chromosome, often visualized as a homogeneously staining region (hsr) or, it may be present as free circular or linear structures, double minutes (Guan *et al.*, 1999; Singer *et al.*, 2000), we examined metaphase spreads from several resistant cell pools using FISH with probes for chromosome 5. We observed rearrangements involving the *DHFR* locus on this chromosome (data not shown), but found no evidence of double minutes, in contrast to a previous report of amplification of *DHFR* in methotrexate resistant HCT116+chr2 cells (Lin *et al.*, 2000). The differing observations may be due to the selection protocols employed or to subtle differences in the HCT116 cells used in the two studies. We note that since HCT116 is MMR deficient, the cells are highly susceptible to continuing mutation in microsatellite sequences. Thus, the genetic backgrounds of HCT116 cell cultures may vary, resulting in different types of genomic alterations providing the most effective drug resistance.

## 5.4 Discussion

Clinical resistance of tumors to methotrexate may be accomplished by a number of routes (Figure 5.6) and frequently involves increased expression of the target gene, *DHFR* (chromosome 5q14) via gene amplification or mutation (Banerjee *et al.*, 2002), down regulation of the reduced folate carrier, *SLC19A1* on chromosome 21q22 (Matherly, 2001), and inefficient polyglutamylation of methotrexate due to decreased activity of folylpolyglutamate synthetase (*FPGS*) on chromosome 9q34 (Rots *et al.*, 1999; Takemura *et al.*, 1999) and up-regulation of  $\gamma$ -glutamate hydrolase (*GGH*) on chromosome 8q12 (Cole *et al.*, 2001). Since methotrexate resistant MMR deficient HCT116 cells displayed few induced copy number changes, we do not know which gene functions have been altered in the resistant cells, but expect that nucleotide level changes may have occurred in certain genes. However, the recurrent copy number changes that did occur were consistent with down regulation of drug influx by loss of one copy of the reduced folate carrier (chromosome 21 loss, 9/29 pools) or up-regulation of *DHFR* (chromosome 5 gain, 4/29 pools). In contrast, we found that the major mechanism of resistance in HCT116+chr3 was up-regulation of *DHFR*, most often by amplification. Thus, MMR proficient and deficient HCT116 cells responded differently to challenge by methotrexate. The observed high frequency of amplifications in MMR proficient cells, and lower frequency of copy number aberrations in MMR deficient cells is consistent with the previously described observations in MMR proficient and deficient cell lines (Figure 2) and with published data on tumors (Muleris *et al.*, 1995; Muleris and Dutrillaux, 1990; Remvikos *et al.*, 1995; Schlegel *et al.*, 1995; Eshleman *et al.*, 1998; Soulie *et al.*, 1999).

We observed two regions of amplification on 5q. One was centered on *DHFR*, the known methotrexate target. The other mapped to a more proximal locus and no previously identified methotrexate resistant gene maps within the  $\sim$ 13 Mb region defined by the amplicon boundaries. However, *CCNBI*, which maps within this region is a candidate, since up-regulation of *CCNBI* due to increased copy number could facilitate progression of MMR proficient cells through G2/M by overcoming a p53-mediated arrest elicited in response to methotrexate-induced DNA damage (Innocente *et al.*, 1999; Taylor and Stark, 2001; Krause *et al.*, 2000; Frouin *et al.*, 2001). Thus, this analysis may have identified another methotrexate resistance gene(s), which is important in the context of functional MMR. However, we cannot rule out the possibility that the copy number profile reflects selection for increased copy number of



**Figure 5.6:** Schematic overview of methotrexate metabolism. Two separate systems, the high affinity folate receptors (*FOLR*, not shown) and the reduced folate carrier (*SLC19A1*) maintain intracellular levels of folate cofactors essential for the synthesis of nucleotides. The reduced folate carrier is the major route by which methotrexate (MTX) is transported into the cell (Matherly, 2001)). Once inside the cell, folylpolyglutamate synthetase (*FPGS*) catalyzes addition of glutamic acid residues (GLU) to methotrexate and folates, causing them to be retained within the cell (Rots *et al.*, 1999; Takemura *et al.*, 1999; Turner *et al.*, 2000). Although glutamylation of methotrexate does not affect affinity for the target enzyme dihydrofolate reductase (*DHFR*), in many cases the modified folate is a better enzyme substrate (Rots *et al.*, 1999; Turner *et al.*, 2000). Inhibition of *DHFR* leads to reduction in deoxythymidine triphosphate and purine nucleotides, causing mis-incorporation of uracil by DNA polymerase under the resulting conditions of limiting nucleotides. Subsequently, DNA strand breaks may result from cycles of removal and re-incorporation of uracil or other mis-incorporated nucleotides due to imbalances in nucleotide pools (Frouin *et al.*, 2001). Methotrexate can be effluxed from cells by energy-dependent pumps (*ABCC1-4*) after removal of glutamic acid residues (Zeng *et al.*, 2001; Chen *et al.*, 2002) by  $\gamma$ -glutamic acid hydrolase (Cole *et al.*, 2001). Clinical and experimental resistance to methotrexate has been reported to involve alterations in the activity of a number of the proteins encoded by these genes. Down regulation of the reduced folate carrier encoded by *SLC19A1* may be accomplished by mutations in the gene, aberrant mRNA splicing, deletions, translocations leading to transcriptional silencing and increased rate of turnover (Matherly, 2001). Modifications of *DHFR* activity in methotrexate resistant cells include up-regulation of expression by increased DNA copy number, promoter mutations and amplification of mutant *DHFR* with low affinity for methotrexate (Banerjee *et al.*, 2002). Inefficient polyglutamylation of methotrexate has also been reported in resistant cells due to low levels of *FPGS* activity (Rots *et al.*, 1999; Takemura *et al.*, 1999) and in some cases high levels of activity of *GGH* (Cole *et al.*, 2001). However, polyglutamylation of folates is necessary for cell proliferation. Some resistant cells have been reported to have high levels of folates, which may not only compensate for inefficient polyglutamylation, but may also contribute to resistance by competition with methotrexate (Takemura *et al.*, 1999). In contrast, the efflux pumps (*ABCC1-4*) are expected to contribute little to resistance in the presence of continuous methotrexate treatment (Zeng *et al.*, 2001; Chen *et al.*, 2002), since polyglutamylated methotrexate cannot be transported, although they may contribute to resistance in the presence of low methotrexate levels. Increased activity could also result in greater efflux of folates, thereby rendering cells more sensitive by reducing folate competition with methotrexate. Chromosomal locations of genes are shown in parentheses.

features of the genome important for *DHFR* amplicon formation or maintenance (e.g. replication origins).

Gene amplification has been studied in vitro in a variety of systems by selection for cells capable of growth in the presence of certain drugs or in mouse models susceptible to gene amplification. These studies indicate that gene amplification is initiated by a DNA double strand break, and that it occurs in cells that are able to progress inappropriately through the cell cycle with this damaged DNA (Kuo *et al.*, 1994; Coquelle *et al.*, 1997; Pipiras *et al.*, 1998; Chernova *et al.*, 1998; Paulson *et al.*, 1998; Zhu *et al.*, 2002). In some cases amplicon boundaries have been mapped to recurrent locations coincident with experimentally induced double strand breaks or common chromosomal fragile sites (Kuo *et al.*, 1994; Coquelle *et al.*, 1998; Hellman *et al.*, 2002), which are 200-300 kb regions often associated with a high frequency of recombinogenic events, including recurrent chromosome aberrations associated with various cancers (Sutherland, 1991; Sutherland *et al.*, 1998; Sutherland and Richards, 1995). In our studies, we found that the amplicons generated in the HCT116+chr3 cells were all unique. Thus, our data provide no evidence that a common initiating site of breakage contributed to their initiation. Instead, it appears more likely that amplification was initiated by random double strand breaks subsequent to mis-incorporation of nucleotides in the presence of methotrexate-induced imbalance in nucleotide pools.

In both MMR proficient and deficient drug resistant cells, we observed a few copy number aberrations, that occurred in only one cell pool and that did not involve genomic regions currently known to harbor resistance genes. These aberrations might be considered bystanders that do not themselves confer resistance. They could be participants in chromosomal rearrangements that resulted in the required alteration of gene expression on the partner chromosome, or non-functional aberrations captured by chance in the resistant cells. On the other hand, they may represent less frequent solutions to overcome inhibition by methotrexate that may be favored in certain circumstances, or they may reveal previously unrecognized mechanisms of drug resistance. For example, in one cell pool, we observed copy number alterations affecting the methotrexate metabolic pathway at three different sites. These cells gained chromosome 5 and lost 21. In addition, they gained a copy of distal 9q including *FPGS*, which is perhaps unexpected, since it is likely to increase efficiency of polyglutamylation and thus retention of methotrexate in the cells. However, this constellation of changes may provide an optimal balance of activities that promote methotrexate resistance through reduced drug influx (loss of 21) and increased copy number of *DHFR*, while at the same time ensuring that sufficient levels of polyglutamated folates are present for cell proliferation (Takemura *et al.*, 1999; Cole *et al.*, 2001). These observations have implications for the interpretation of aberrations in tumor genomes, since they suggest that most genomic alterations are likely to contribute to the neoplastic process, and only rarely are they bystanders.

Methotrexate resistance appears to be achieved by a multi-step selection process, as evidenced by the presence of alterations in several genes, as well as the chromosome 5 amplicons in MMR proficient cells. Thus, significant selection is operating on subtle adventitiously arising genomic alterations, each one of which may slightly enhance the viability of the cells. A similar accumulation of subtle events is likely to be involved in tumorigenesis. Models of human cancer highlight the importance of up-regulation of growth promoting genes together with down regulation of apoptotic signaling. By altering the expression of a small number of carefully chosen genes it is possible to experimentally generate transformed cells or tumors that display almost all aspects of the tumor phenotype, including invasion, metastasis and angiogenesis (Zimonjic *et al.*, 2001; Pelengaris *et al.*, 2002). One feature of human cancers

that is not recapitulated in these models is the frequent and varied alterations that occur in the genomes of cells in solid tumors. This apparent increased complexity is likely to reflect the contribution of genome rearrangements to the evolutionary selection for incremental changes in gene expression that promote cell proliferation and survival in human tumors, in contrast to the well-considered choices made when manipulating mouse models and cultured cells. Thus, as cells tiptoe toward malignancy, DNA copy number changes provide one mechanism by which cells can seek the solutions that are just right for competitive advantage, with the result that their genomes are shaped both by selection for beneficial changes in gene expression and by failures in maintenance of genomic integrity that allow certain types of aberration to occur with higher probability.

## 5.5 Materials and Methods

### 5.5.1 Cells

We obtained DU145, SKOV3, HCT116, AN3CA and SKUT1 from the UCSF Cell Culture Facility, and Jurkat, LoVo, DLD-1, LS174T and SW48 from the American Type Culture Collection (ATCC, Manassas, VA). Dr. C. R. Boland generously provided the mismatch repair proficient cell line HCT116+chr3 (Koi *et al.*, 1994; Hawn *et al.*, 1995), containing a wild type copy of MLH1. We confirmed the presence of mlh1 protein by western blot. We did not use a companion cell line, HCT116+chr2, which others have used (Lin *et al.*, 2000), because our array CGH analysis of the cell line revealed a single copy gain of chromosome 21 and no evidence of additional chromosome 2 material. The complete data set is available at: <http://cc.ucsf.edu/albertson/public/>.

### 5.5.2 Cell culture and selection of resistant cell pools

We maintained HCT116 and HCT116+chr3 in DMEM supplemented with 4.5 g/L glucose, 0.584 g/L L-glutamine, 3.7 g/L NaHCO<sub>3</sub>, 10% dialyzed fetal bovine serum (HyClone, UT), 1x penicillin and 1x non-essential amino acids. We supplemented the media for HCT116+chr3 with 400  $\mu$ g/mL geneticin disulfate. We dissolved methotrexate (Sigma) in 0.1 M NaOH and diluted it to 100 or 500  $\mu$ M in Hank's BSS. We challenged cells with methotrexate using five different protocols for setting up the cultures in order to increase the probability that resistant cell pools arose as independent events. For protocol 1, we seeded  $4 \times 10^5$  cells in 100 mm plates (pools O1-O5, P2) or T75 flasks (protocol 2, pools 4-11, E1-4, K1, K5, L1, M1-3, Q1, Q2, R1 and R2) and allowed cells to recover overnight before changing to fresh medium containing 25 nM methotrexate. We also seeded  $3 \times 10^3$  cells (protocol 3, pools A-C) or 100 cells (protocol 4, pools I2 and I7) in each well of multi-well plates, expanded the cells in each well to  $4 \times 10^5$  cells and then transferred them to T75 flasks. We allowed the cells to recover overnight before changing to medium containing 25 nM methotrexate. For protocol 5 (pools 1-3), we seeded  $10^4$  cells in each of 3 wells of a 24 well plate and allowed the cells to recover overnight before changing to medium containing 25 nM methotrexate. In all cases, we changed the medium every 2-3 days until resistant colonies arose (2-3 weeks). We pooled the resistant colonies from each flask or plate or isolated a single resistant colony (pools O1-5 and P2) and expanded them in media containing 25 nM methotrexate. Resistant colonies arose in HCT116 cultures at  $\sim 50$  times higher frequency than in HCT116+chr3 cultures ( $142 \pm 64/400,000$  cells compared to  $2.8 \pm 1.4/400,000$  cells). In several cases, we

compared the copy number profiles obtained from individual colonies isolated from a flask with the pool of cells from the same flask. Since the same genomic aberrations were observed in both, indicating a common progenitor, the analyses of cell pools were considered equivalent to individual isolated clones. We verified by array CGH that the copy number profiles of genomic DNA from cells cultured in parallel, but not treated with methotrexate, remained stable. In addition, we analyzed *TP53* mutation status, by sequencing exons 2-11 in untreated HCT116 and HCT116+chr3 cells used to seed plates for selection of resistant pools O and P, resistant HCT116 cell pools O1 and O4 and HCT116+chr3 pools M1-M3, P2, R1 and R2. In all cases, the sequence was wild type.

### 5.5.3 DNA isolation

We incubated cells from a T75 flask overnight at 55°C in a 3 ml solution containing 0.01 M Tris, pH 7.5, 0.001 M EDTA, pH 8, 0.5% SDS and 0.1  $\mu\text{g}/\mu\text{l}$  proteinase K. We precipitated the DNA with ethanol, recovered it by spooling and then dissolved it in 300  $\mu\text{l}$  H<sub>2</sub>O.

### 5.5.4 DNA labeling

We labeled genomic DNA by random priming in a 50  $\mu\text{l}$  reaction by combining 300 ng genomic DNA, 1x random primers solution (BioPrime DNA Labeling System, Gibco BRL), 1x dNTP solution (0.2 mM each of dATP, dTTP, dGTP, 0.05 mM dCTP, 1 mM Tris base (pH 7.6), 0.1 mM EDTA), 40  $\mu\text{M}$  Cy3 or Cy5-dCTP (Amersham) and 40 U Klenow fragment (BioPrime DNA Labeling System, Gibco BRL). We incubated the DNA and the random primers in a volume of 42  $\mu\text{l}$  at 100°C for 10-15 min, placed the reaction on ice and added the dNTP solution, labeled nucleotides and Klenow fragment. We incubated the random priming reaction overnight at 37°C. We removed unincorporated nucleotides using a Sephadex spin-column.

### 5.5.5 BAC arrays

We used two versions of the whole genome human BAC arrays described previously (Snijders *et al.*, 2001). Each array is comprised of 2464 clones printed in triplicate. HumArray1.14 (Snijders *et al.*, 2001) contains 2277 clones that map to a single site in the genome by FISH. HumArray2.0 is the next generation genome array and contains 2443/2464 clones that can be mapped onto the genome sequence. To generate HumArray2.0, we removed multi-site clones and clones that mapped improperly from HumArray1.14 and added others to improve coverage. We assembled an array providing higher density coverage of the *DHFR* region at 5q14 by arraying an additional 10 BACs from the region. We obtained a total of 13 BACs from the RP11 library from Dr. N. Nowak (Roswell Park Cancer Institute) and from the CTC and CTD libraries from Invitrogen. We mapped each BAC onto metaphase chromosomes by FISH to confirm that it mapped to a single location at 5q14 and excluded 3 BACs because they mapped to multiple sites in the genome. We also included 200 clones distributed across the genome on the array. We used these clones for normalization, so that hybridization ratios from the *DHFR* array could be compared to the genome-wide data obtained on HumArray2.0.

### 5.5.6 Array CGH hybridization

We combined and precipitated 50  $\mu$ l Cy3 labeled test DNA, 50  $\mu$ l Cy5 labeled reference DNA and 100  $\mu$ g human Cot-1 DNA by adding 0.1 volumes of 3 M sodium acetate (pH 5.2) and 2 volumes of ice-cold ethanol. We collected the precipitate by centrifugation and dissolved the pellet in a final hybridization mixture containing 50% formamide, 2xSSC and 4% SDS. We denatured the hybridization mixture at 75°C for 10-15 min and then incubated the samples at 37°C for 1 hour to block repetitive DNA sequences. We applied a ring of rubber cement closely around the array to form a well, into which we added 50  $\mu$ l of a pre-incubation solution containing 250  $\mu$ g salmon sperm DNA, 50% formamide, 2xSSC and 4% SDS. After a 30 min incubation at room temperature, we removed  $\sim$ 30  $\mu$ l of the pre-incubation solution and added the denatured and re-annealed hybridization mixture. We placed a silicone gasket around each hybridization area, placed a microscope slide on top of the gasket, which was thick enough to prevent the microscope slide from contacting the hybridization solution and clamped the whole unit together using binder clips to create an air tight environment. We carried out the hybridization to the arrays for  $\sim$ 48 hrs on a slowly rocking table at 37°C. After disassembling the hybridization chamber, we rinsed off the excess hybridization fluid with PN buffer (PN: 0.1 M sodium phosphate, 0.1% nonidet P40, pH 8), then washed the arrays once in a solution containing 50% formamide in 2xSSC for 15 min. at 45°C and finally in PN buffer at room temperature for 15 min. Subsequently, we removed the ring of rubber cement, drained excess liquid from the arrays, mounted them in a solution containing 90% glycerol, 10% PBS and 1  $\mu$ M DAPI, and sealed them with a cover slip.

### 5.5.7 Imaging and analysis

We acquired 16 bit 1024x1024 pixel DAPI, Cy3 and Cy5 images using a custom built CCD camera system (Pinkel *et al.*, 1998). We used "UCSF SPOT" software (Jain *et al.*, 2001) to automatically segment the spots based on the DAPI images, perform local background correction and to calculate various measurement parameters, including  $\log_2$ ratios of the total integrated Cy3 and Cy5 intensities for each spot. We used a second custom program SPROC to associate clone identities and a mapping information file with each spot so that the data could be plotted relative to the position of the BACs on the August, 2001 freeze of the draft human genome sequence (<http://genome.ucsc.edu>). SPROC also implements a filtering procedure to reject data based on a number of criteria, including low reference/DAPI signal intensity and low correlation of the Cy3 and Cy5 intensities with a spot. The SPROC output consists of averaged ratios of the triplicate spots for each clone, standard deviations of the triplicates and plotting position for each clone on the array, as well as other clone information stored in the database, such as STS content. We edited the data files to remove ratios on clones for which only one of the triplicates remained after SPROC analysis and/or the standard deviation of the  $\log_2$ ratios of the triplicates was  $> 0.2$ .

### 5.5.8 Statistical methods

For a given sample, we identified copy number transitions on the chromosomes by applying unsupervised Hidden Markov Model methodology (Rabiner, 1989) to discover spatially coherent homogeneous groups of clones with the same underlying copy number. We merged adjacent stretches consisting of more than two contiguous clones assigned to the same state if the medians of these stretches were not sufficiently different (the absolute difference less



than 0.2) or if the medians of both stretches were close enough to 0 in their absolute values ( $<0.25$ ). If a chromosome contained  $N$  stretches, we recorded  $(N-1)$  transitions and arbitrarily placed the transition points at the last clone of the first of the two adjacent stretches. We identified whole chromosomes as gained or lost depending on the value of their median, if they contained no transitions and their absolute median was greater than 0.25. We recorded single clones or pairs of clones as focal aberrations if their values were further away from the median of the corresponding stretch than 4 times the median absolute deviation of that stretch. Finally, we counted a clone or a group of clones as a focal amplification if (a) the values of the  $\log_2$ ratios were greater than 1 and (b) the region containing clones with elevated ratios ( $\log_2\text{ratio} > 0.7$ ) only spanned  $<10$  Mb. We used the Wilcoxon rank sum test to assess the significance of the differences between groups.

## **5.6 Acknowledgements**

We thank M. Koi and C. R. Boland for providing HCT116+chr3 cells, N. Nowak and J. Conroy for providing clones from the RP11 library and T. Paulson and G. Wahl for sharing protocols and cells. We thank S. Blackwood, N. Brown, G. Hamilton, B. Huey, S. Kwek and A. Matthäi for their participation in discussions and array production, and K. Kimura and the UCSF Cancer Center Bioinformatics Core for microarray database support. This work was supported by NIH grants CA94407 and CA83040.



## Chapter 6

# Genome-Wide Array Based Comparative Genomic Hybridization Reveals Genetic Homogeneity and Frequent Copy Number Increases Encompassing *CCNE1* in Fallopian Tube Carcinoma

Antoine M. Snijders<sup>1,2</sup>, Marlies E. Nowee<sup>3</sup>, Jane Fridlyand<sup>1,2</sup>, Jurgen M. J. Piek<sup>3</sup>,  
Josephine C. Dorsman<sup>3</sup>, Ajay N. Jain<sup>1,2,4</sup>, Daniel Pinkel<sup>2,4</sup>, Paul J. van Diest<sup>5</sup>,  
René H. M. Verheijen<sup>3</sup> and Donna G. Albertson<sup>1,2,4</sup>

<sup>1</sup>Cancer Research Institute, University of California San Francisco, San Francisco CA, USA

<sup>2</sup>University of California San Francisco Comprehensive Cancer Center, San Francisco,  
San Francisco CA, USA

<sup>3</sup>Department of Obstetrics and Gynaecology, VU University Medical Center, Amsterdam,  
The Netherlands

<sup>4</sup>Department of Laboratory Medicine, University of California San Francisco, San Francisco  
CA, USA

<sup>5</sup>Department of Pathology, VU University Medical Center, Amsterdam, The Netherlands

Oncogene 22, 4281-4286, 2003

## 6.1 Abstract

Fallopian tube carcinoma (FTC) is a rare, poorly studied and aggressive cancer, associated with poor survival. Since tumorigenesis is related to acquisition of genetic changes, we used genome-wide array CGH to analyze copy number aberrations occurring in FTC in order to obtain a better understanding of FTC carcinogenesis and to identify prognostic events and targets for therapy. We used arrays of 2464 genomic clones, providing ~1.4 Mb resolution across the genome to quantitatively map genomic DNA copy number aberrations from fourteen FTC onto the human genome sequence. All tumors showed a high frequency of copy number aberrations with recurrent gains on 3q, 6p, 7q, 8q, 12p, 17q, 19 and 20q, and losses involving chromosomes 4, 5q, 8p, 16q, 17p, 18q and X. Recurrent regions of amplification included 1p34, 8p11-q11, 8q24, 12p, 17p13, 17q12-q21, 19p13, 19q12-q13 and 19q13. Candidate, known oncogenes mapping to these amplicons included *CMYC* (8q24), *CCNE1* (19q12-q21) and *AKT2* (19q13), whereas *PIK3CA* and *KRAS*, previously suggested to be candidate driver genes for amplification mapped outside copy number maxima on 3q and 12p, respectively. The FTC were remarkably homogeneous, with some recurrent aberrations occurring in more than 70% of samples, which suggests a stereotyped pattern of tumor evolution.

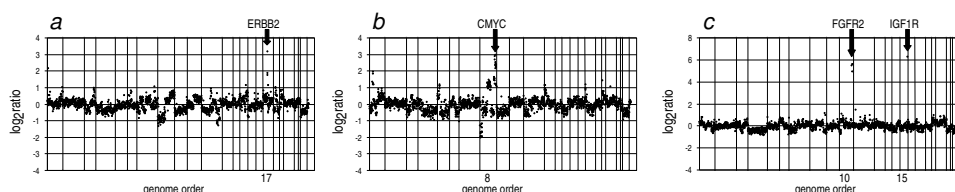
## 6.2 Introduction

Fallopian tube carcinoma (FTC) is an aggressive form of cancer in women that is associated with a poor survival. Previous studies have reported that FTC show an aneuploid DNA content, overexpression of *TP53*, *ERBB2* and *CMYC* (Hellström *et al.*, 1994; Chung *et al.*, 2000), as well as a high frequency of *KRAS* mutations (Mizuuchi *et al.*, 1995). Deficiency for *BRCA1* may also contribute to the genesis of FTC, since the incidence of FTC is higher in families carrying *BRCA1* mutations (Zweemer *et al.*, 2000; Aziz *et al.*, 2001; Leeper *et al.*, 2002) and dysplasia is a frequent finding in prophylactically removed Fallopian tubes from *BRCA1* carriers (Piek *et al.*, 2001). However, the application of chromosome comparative genomic hybridization (CGH) to the study of FTC revealed that these tumors are characterized by a high level of copy number aberrations with frequent amplification of 3q (Heselmeyer *et al.*, 1998; Pere *et al.*, 1998). Thus, there appear to be many genes, in addition to the commonly studied oncogenes and tumor suppressors that contribute to the genesis of FTC. Detecting and quantifying these copy number changes may lead to a better understanding of FTC carcinogenesis and might identify prognostic events and targets for the choice or design of therapy.

## 6.3 Results

Here, we report on the application of array CGH to further refine regions of recurrent aberration in FTC. Array CGH provides a means to quantitatively detect genomic DNA copy number aberrations and map them directly onto the sequence of the human genome (Solinas-Toldo *et al.*, 1997; Pinkel *et al.*, 1998; Pollack *et al.*, 1999; Snijders *et al.*, 2001). We used an array comprised of 2464 genomic clones, each mapped to the August 2001 freeze of the human genome sequence, to determine the copy number profiles of 14 FTC. The tumors ranged

in FIGO stage from Ia to IV and included one grade 1, seven grade 2 and six grade 3 tumors. Patient age at diagnosis ranged from 46 - 82 years, and one patient met the criteria for family cancer syndrome. Copy number profiles for three tumors are shown in Figure 6.1 and illustrate the high frequency of copy number alterations in these tumors in agreement with the limited chromosome CGH data reported previously (Heselmeyer *et al.*, 1998; Pere *et al.*, 1998).



**Figure 6.1:** Genome-wide DNA copy number analysis of three Fallopian tube carcinoma. Array CGH was carried out using arrays of 2464 BAC clones each printed in triplicate (HumArray2.0) according to published protocols (Snijders *et al.*, 2001, 2003). The majority of the BACs used for the array were obtained from FISH verified clone sets (Knight *et al.*, 2000; Cheung *et al.*, 2001). The array has been extensively validated for its capability to reliably detect and map changes from single copy in a diploid background, as well as high level amplifications (Snijders *et al.*, 2001). Tumor DNAs were extracted from 20 microdissected consecutive 10  $\mu$ m formaldehyde fixed paraffin embedded sections (Zweemer *et al.*, 2001) using a QIAmp DNA mini kit following a modification of the manufacturer's protocol (Weiss *et al.*, 1999). The tumor genomic DNA and normal female reference DNA (300 ng each) were labeled by random priming in separate 50  $\mu$ l reactions to incorporate Cy3 and Cy5, respectively. The test and reference DNAs together with 100  $\mu$ g human Cot-1 DNA were hybridized to the BAC arrays for  $\sim$ 48 hrs at 37°C. After post-hybridization washes, the arrays were mounted in a solution containing 90% glycerol, 10% PBS and 1  $\mu$ M DAPI and sealed with a cover slip. A custom built CCD camera system was used to acquire 16 bit 1024x1024 pixel DAPI, Cy3 and Cy5 images (Pinkel *et al.*, 1998). Image analysis was carried out using UCSF SPOT software (Jain *et al.*, 2001). The log<sub>2</sub>ratio of the total integrated Cy3 and Cy5 intensities for each spot after background subtraction was calculated, normalized to the median log<sub>2</sub>ratio of all the clones on the array and the average of the triplicates calculated using a second custom program, SPROC. Automatic data filtering to reject data points based on low DAPI intensity, low correlation between Cy3 and Cy5 within each segmented spot and low reference/DAPI signal intensity was also carried out using SPROC. Data files were subsequently manually edited by rejecting clones for which only one spot of the triplicate survived after SPROC analysis and for which the standard deviation of the log<sub>2</sub>ratio of the triplicate was  $>0.2$ . For each tumor, the data are plotted as the mean log<sub>2</sub>ratio of the triplicate spots for each clone normalized to the genome median log<sub>2</sub>ratio. The clones are ordered by position in the genome (UCSC draft genome sequence, August 2001 freeze) beginning at 1p and ending with Xq. The borders between chromosomes are indicated with vertical bars. All tumors show many low-level DNA copy number aberrations as well as high-level amplifications of parts of the genome. High-level amplifications can be seen for *ERBB2* located on chromosome 17 (a), *CMYC* located on chromosome 8 (b) and for *FGFR2* and *IGF1R* on chromosome 10 and 15, respectively (c).

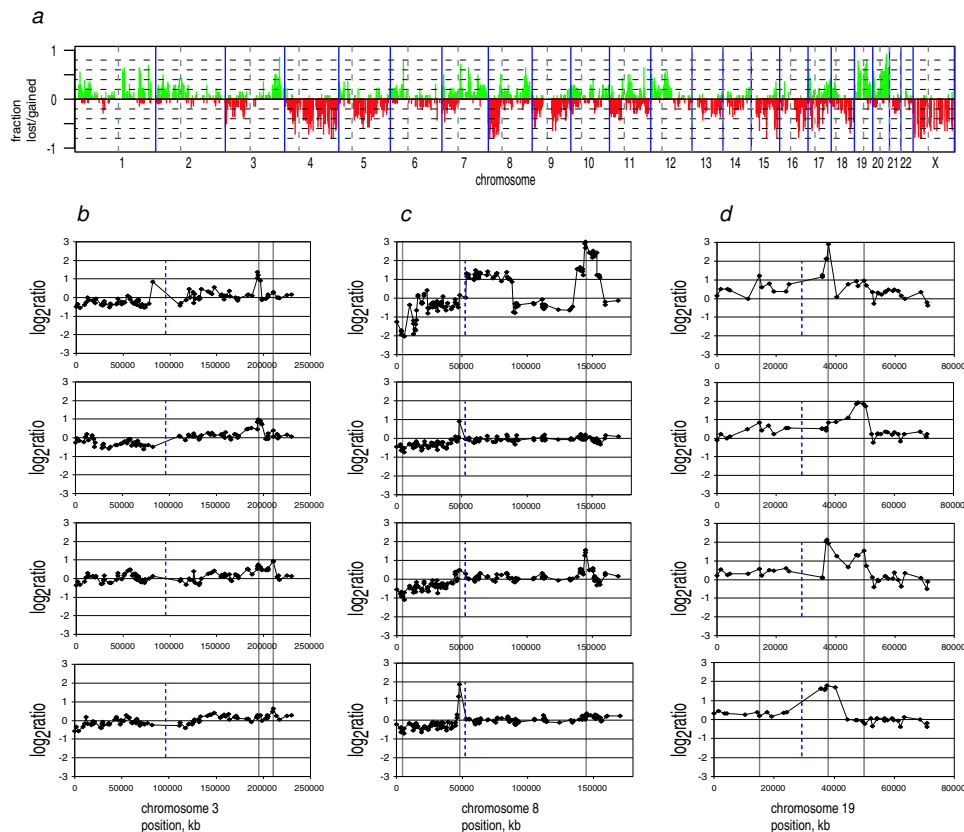
Using previously described statistical methods (Snijders *et al.*, 2003), we determined the frequency of the different types of copy number aberrations distinguished by array CGH,

including low level copy number alterations affecting whole chromosomes or portions of chromosomes, and more focal aberrations such as deletions and high level copy number increases, or amplifications. We observed an average of 94 aberrations per tumor (range 37-146). Copy number alterations involving whole chromosome gains and losses were rare. On the other hand, copy number alterations affecting chromosome arms or portions of arms, which we scored as transitions from one copy number to another along a chromosome were frequent (average 55/tumor, range 22 to 85 per tumor). There were no significant correlations of copy number aberrations with stage or grade. We observed recurrent regions of aberration as shown by the frequency plot of gains or losses recorded on each clone (Figure 6.2a). Many of these regions were found previously by chromosome CGH, including recurrent copy number gains on 3q, 6p, 7q, 8q, 12p, 17q and 20q, and losses involving chromosomes 4, 5q, 8p, 16q, 17p and 18q. However, the application of array CGH mapped these regions with higher resolution and also found regions of chromosome 19 and X that were consistently gained or amplified (14/14 tumors) and lost (13/14 tumors), respectively. Regions of amplification were also observed, a number of which had previously not been reported (Table 6.1). Several of these regions would have been difficult to detect by chromosome CGH as they map near the centromeres or telomeres of the chromosomes. These regions, in addition to chromosome 19, often display copy number artifacts by chromosome CGH (Kallioniemi *et al.*, 1994; Karhu *et al.*, 1997), which are not encountered in array CGH (Snijders *et al.*, 2001).

**Table 6.1:** Recurrent regions of amplification found by array CGH in 14 Fallopian tube adenocarcinomas

<i>Chromosome location</i>	<i>Flanking proximal clone</i>	<i>Flanking distal Flanking distal</i>	<i>Number of cases</i>	<i>Candidate genes</i>
1p34	RP11-219O7	RP11-88O2	2	<i>MYCL1, MYCBP</i>
8p11	RP11-48D21	RP11-73M19	2	<i>PLAT, POLB, ANK1</i>
8q24	RP11-128G18	RP11-227F7	3	<i>CMYC</i>
17p13	GS1-68F18	RP11-4F24	3	<i>CRK</i>
17q12-q21	RP11-58O8	CTD-2094C6	2	<i>ERBB2</i>
19p13	RP11-84C17	RP11-107O2	3	<i>JUNB, RAD23A, BRG1, RAB3D, RAB11D</i>
19q12-q13	RP11-152P7	RP11-147D7	4	<i>CCNE1</i>
19q13	RP11-91H20	RP11-133A7	2	<i>AKT2</i>

Gain or amplification of the distal part of chromosome 3q, centering on 3q26, has been reported in many tumors. Here, we resolved two regions of gain on distal 3q, including one centered on 3q25-26 (7/14 tumors) and one centered on 3q26-27 (10/14 tumors). Candidate genes that map to the 3q25-26 region, which is flanked by RP11-198G24 and RP11-172G5, include *TERC*, which has been suggested as a driver gene for this amplification (Heselmeyer *et al.*, 1998), as well as *TNFSF10*, *EVII* and *SKIL*, which are also good candidates. The more distal region of gain at 3q26-27 is flanked by RP11-118F4 and CTD-2091K6. Candidate genes in this region include *THPO* and *MAP3K13*. The oncogene, *PIK3CA*, which has been reported to be at increased copy number in many ovarian tumors maps in between these two regions of frequent copy number gain/amplification on distal 3q. If driver genes for amplification map to the copy number maximum in regions of amplification (Albertson



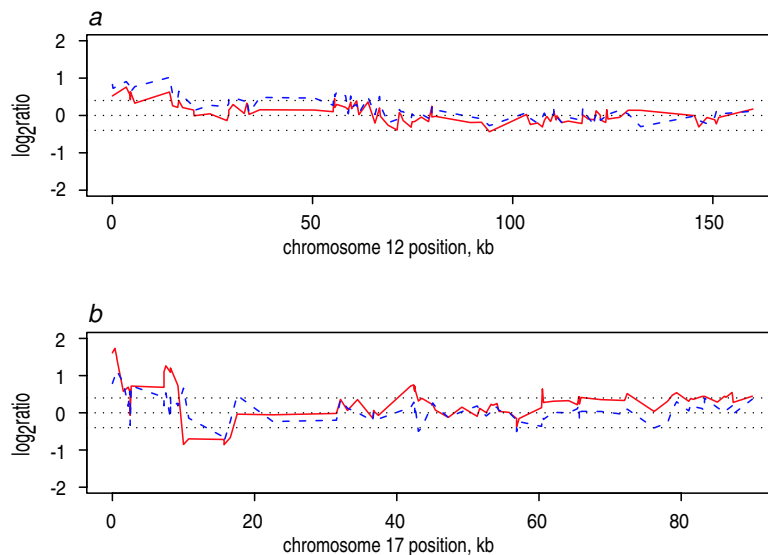
**Figure 6.2:** DNA copy number analysis of 14 FTC. (a) Frequency of gains (ranging from 0 to 1, green bars) with  $\log_2\text{ratio} > 0.4$  and frequency of losses (ranging from 0 to -1, red bars) with  $\log_2\text{ratio} < -0.4$  for each clone in 14 FTC. Clones are plotted in order of genome position (UCSC draft genomic sequence, August 2001 freeze) with vertical bars indicating chromosome boundaries and dotted vertical lines indicating the position of the centromere for each chromosome. (b-d) Copy number profiles showing recurrent aberrations of chromosomes 3 (b), 8 (c) and 19 (d). Normalized  $\log_2\text{ratio}$ 's for clones are plotted based on their relative position on the UCSC draft genomic sequence (August 2001 freeze). The location of the centromere is indicated by a dashed vertical bar. Recurrent regions of copy number gain/amplification or loss are indicated with a vertical bar spanning all four plots for each chromosome.

*et al.*, 2000), then these observations suggest that *PIK3CA* is not the driver gene for either recurrent aberration. However, increased copy number of *PIK3CA* was observed in six FTC and thus increased copy number of *PIK3CA* may still contribute to the genesis of FTC. Loss of 8p has frequently been observed in many solid tumors. Here, we found 8p loss in all FTC tumor samples. The minimal region of recurrent loss was located at 8p23 in 12/14 tumors and is flanked by RP11-82K8 and RP11-140K14. The cub and sushi multiple domains-1 gene, *CSMD1* is currently the only named gene in this region (UCSC June 2002 freeze). Although,

the region is frequently deleted in a number of different tumor types, little is known about the function of this gene (Sun *et al.*, 2001). We also found a previously unreported region of gain (5/14 tumors) or amplification (2/14 tumors) mapping at proximal 8p. The region is flanked by RP11-48D21 and RP11-73M19 and does not include *FGFR1* (Figure 6.2c). Several candidate oncogenes mapping to this region include *PLAT*, *POLB* and *ANK1*. A high frequency of *KRAS* mutations has been previously reported in FTC (Mizuuchi *et al.*, 1995). Although we find chromosome 12p to be gained or amplified in at least six FTC, *KRAS*, which is included in 2/6 copy number gains, maps outside of the minimal recurrent region of copy number gain, which is flanked by GS1-124K20 and RP11-24N12. Candidate oncogenes mapping within the region include *RBBP2*, *RAD52* and *WNT5B*. Three regions of chromosome 19 were gained or amplified with one or more regions being affected in all 14 tumors. Aberrations at 19p13-p12 were observed in 12/14 tumors. This minimal region is flanked by RP11-84C17 and RP11-107O2 (Figure 6.2d). Several cancer related genes map to this region, including *JUNB*, *RAD23A*, *BRG1*, *RAB3D* and *RAB11B*. The second region of frequent copy number increase is located at chromosome 19q13 (Figure 2d). This region is gained in 9/14 FTC and amplified in 2/14 FTC. The minimal region of amplification is flanked by RP11-92J4 and RP11-208I3 and includes the oncogene *AKT2*, coding for a serine-threonine protein kinase involved in increased survival and proliferation. Previous studies have shown that *AKT2* is amplified/overexpressed (Cheng *et al.*, 1992; Bellacosa *et al.*, 1995) or activated in ovarian cancers (Yuan *et al.*, 2000). However, *AKT2* is not included in 3/9 copy number gains suggesting that other genes in this region contribute to FTC carcinogenesis. Amplification of the third region on chromosome 19 at 19q12-13, flanked by RP11-152P7 and RP11-147D7 was observed in four FTC. The region was also gained in an additional six FTC. The copy number maximum of this amplicon encompasses *CCNE1*. Previous studies have found *CCNE1* amplifications in ovarian, colon, gastric, esophageal and urinary bladder cancer (Marone *et al.*, 1998; Kitahara *et al.*, 1995; Akama *et al.*, 1995; Lin *et al.*, 2000; Richter *et al.*, 2000). Up-regulation of *CCNE1* may contribute to tumorigenesis by accelerating progression through G1/S, although other oncogenic functions have been described, in addition to its ability to promote cell cycle progression (Geisen and Möröy, 2002), including, for example, malignant transformation in cooperation with *HRAS* and *CMYC* (Haas *et al.*, 1997) and genomic instability (Spruck *et al.*, 1999). Losses of the X chromosome or parts thereof were observed in 13/14 FTC and will result in different functional consequences depending on whether the active or inactive X is lost. If the active X is lost, the cells will be effectively homozygous null for genes subject to X inactivation, while loss of the inactive X should result in altered expression of fewer genes (those that escape X inactivation). A tendency toward loss of the late replicating X has been reported previously in solid tumors in females (Dutrillaux *et al.*, 1986) and thus, its loss may provide a proliferative advantage. The observed high frequency of recurrence of aberrations, some occurring in 70% or more of the cases indicated that the Fallopian tube cancers were remarkably homogeneous in their copy number profiles (Figure 6.3).

A comparison of the copy number aberrations across the set of 14 tumors revealed that the data set was highly correlated (average Spearman rank correlations of the pairs of tumors is 0.47). We have previously reported that specific defects in pathways that normally maintain genomic stability give rise to certain types and frequencies of copy number aberrations (Snijders *et al.*, 2003). The data reported here suggest a limited number of routes lead to FTC pathogenesis. It has been noted previously that serous carcinoma of the Fallopian tubes, uterus and ovary are similar histologically and in their clinical behavior, including invasion





**Figure 6.3:** Genetic homogeneity of FTC. Comparison of copy number profiles of chromosomes 12 (a) and 17 (b) from two pairs of FTC (one is shown as a solid line and the other as a dashed line). Data are plotted as in Figure 2d. The dotted lines indicate the  $\log_2\text{ratio} = \pm 0.4$  cutoff used for the aberration frequency plot in Figure 6.2a.

and poor prognosis (Pere *et al.*, 1998). They also share a number of genetic features including a high frequency of *TP53* overexpression (Hellström *et al.*, 1994; Caduff *et al.*, 1998), amplification and overexpression of certain oncogenes including *ERBB2* and *CMYC* in all three types (Monk *et al.*, 1994; Wang *et al.*, 1999; Chung *et al.*, 2000; Manavi *et al.*, 1998) and *AKT2* in ovarian (Yuan *et al.*, 2000) and Fallopian tube cancers as shown here. They also have similar chromosome CGH profiles (Pere *et al.*, 1998). These similarities may reflect the fact that the Fallopian tubes, ovarian surface epithelium and uterus develop from the Mullerian ducts and thus cells from this lineage may tiptoe towards cancer via similar pathways. Further insights await higher resolution analysis of the genome-wide aberrations in these other tumor types by array CGH.

## 6.4 Acknowledgements

We thank S. Blackwood, N. Brown, G. Hamilton, B. Huey, S. Kwek, A. Matthäi and R. Segreaves for their participation in discussions and array production, and K. Kimura and the UCSF Cancer Center Bioinformatics Core for microarray database support. This work was supported by NIH grants CA94407 and CA83040, by grant 901-02-241 from NOW and by a grant from the Biocare Foundation.



## Chapter 7

# Demonstration of Oral Squamous Cell Carcinoma Second Primary Clonality Using Genome Wide Array Comparative Genomic Hybridization

Brian L. Schmidt <sup>1\*</sup>, Antoine M. Snijders <sup>2\*</sup>, Jane Fridlyand <sup>3</sup>, Nusi Dekker <sup>4</sup>, Richard C. K. Jordan <sup>4,5</sup>, Sol Silverman Jr. <sup>4</sup> and Donna G. Albertson <sup>2,6</sup>

\*These authors contributed equally.

<sup>1</sup>Department of Oral and Maxillofacial Surgery, University of California San Francisco, San Francisco CA, USA

<sup>2</sup>Cancer Research Institute, University of California San Francisco, San Francisco CA, USA

<sup>3</sup>Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco CA, USA

<sup>4</sup>Department of Stomatology, University of California San Francisco, San Francisco CA, USA

<sup>5</sup>Department of Pathology, University of California San Francisco, San Francisco CA, USA

<sup>6</sup>Department of Laboratory Medicine, University of California San Francisco, San Francisco CA, USA

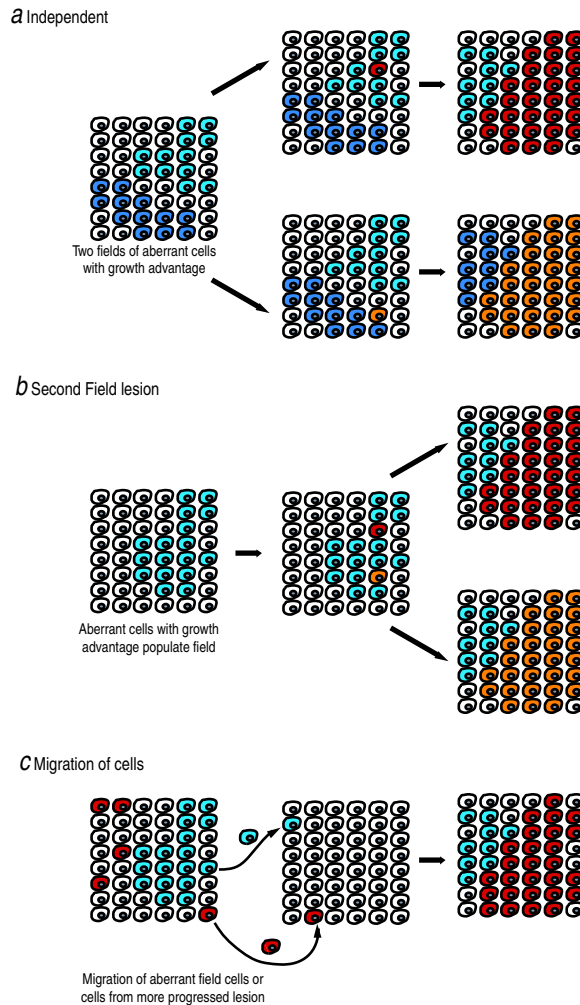
Submitted

## 7.1 Introduction

Management of head and neck cancer is complicated by local recurrence and the development of second primaries (Sudbo *et al.*, 2004). Following diagnosis of one primary oral squamous cell carcinoma (SCC) the annual rate for development of a second primary SCC is 4-7% (Batsakis, 1984; Cooper *et al.*, 1989; Sturgis and Miller, 1995) and the 5-year survival rate is reduced to 25%. Second primary oral SCCs are hypothesized to be; (a) independent tumors that developed from unrelated cells, possibly within a field of cells predisposed due to environmental agents such as tobacco smoke (field cancerization) (Slaughter *et al.*, 1953), (b) "second field tumors" (Tabor *et al.*, 2002) or "clonal cancers" (Partridge *et al.*, 2001) that developed from a field or patch of aberrant cells that had spread to encompass large areas of the oral cavity, or (c) tumors established by cells sloughed or migrating from a previous tumor (Figure 7.1).

Since tumor genomes typically acquire multiple alterations, the relatedness of multiple tumors can be evaluated based on whether or not they share particular alterations (Albertson *et al.*, 2003). Thus, tumors arising by the first mechanism would be likely to share only those genetic aberrations that are most common to oral SCCs, whereas those arising from the second and third mechanisms are likely to share not only aberrations frequently associated with oral SCCs, but also those that were present in the field (second primary field tumors) or in the tumor from which a cell had migrated. Previous studies investigating the possible origin of recurrent tumors have used a limited number of markers mostly representing commonly occurring aberrations to establish whether the tumors are genetically related, and have led to both conclusions - that tumors arose independently or were clonally related (Partridge *et al.*, 2001; Jang *et al.*, 2001; Braakhuis *et al.*, 2002; Ai *et al.*, 2001; Wolf *et al.*, 2004; Waridel *et al.*, 1997; Carey, 1996). Since the small number of markers assayed in these previous studies may have failed to reveal clonality, we used array comparative genomic hybridization (array CGH) to perform a genome-wide analysis of multiple oral SCC, clinically classified as second primary tumors because they occurred at a site >2 cm distant from the initial tumor and/or more than 3 years after the initial tumor (almost 17 years in one patient). In array CGH, total genomic DNAs from a "test" and a "reference" cell population are labeled with different fluorochromes and hybridized to arrays of genomic clones that have been mapped to specific locations in the genome. The ratios of the fluorescence intensities of the two fluorochromes are approximately proportional to the ratio of the copy number of the corresponding DNA sequences in the test and reference genomes (Pinkel *et al.*, 1998; Snijders *et al.*, 2001) and provide a distinctive copy number profile across the genome for each tumor (Figure 7.2).

We observed "signature" copy number alterations that were present in each of the multiple tumors from an individual patient, indicating that the tumors are clonally related. These observations, which are consistent with the tumors originating as second field tumors or by seeding from tumor cells from an earlier tumor, highlight the utility of molecular analysis for definition of second primary tumors in patients with multiple oral SCCs and indicate the utility of molecular analysis for identification and management of patients at risk for second primaries.

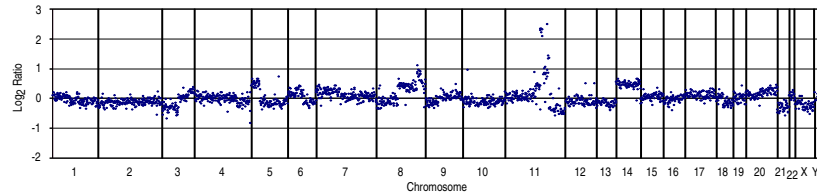


**Figure 7.1:** Models for origin of second tumors. **a.** Independent. Second primaries might arise independently from genetically different cells from different fields (light blue and dark blue cells). **b.** Second field tumors or clonal cancerization. Clonally related tumors could arise in a single field of aberrant cells (blue cells). **c.** Migration. Premalignant cells from a field of aberrant cells (blue) or cells shed from an earlier tumor (red) could populate a distant site and give rise to genetically related multifocal second primary tumors.

## 7.2 Case Reports

### 7.2.1 Patient 1

The patient is a 67-year-old woman with no history of tobacco or alcohol use who presented with a left posterior mandibular gingival ulcerative lesion (Figure 7.3, tumor A). A biopsy demonstrated SCC. She was treated with a resection of gingiva, teeth and alveolar bone. All

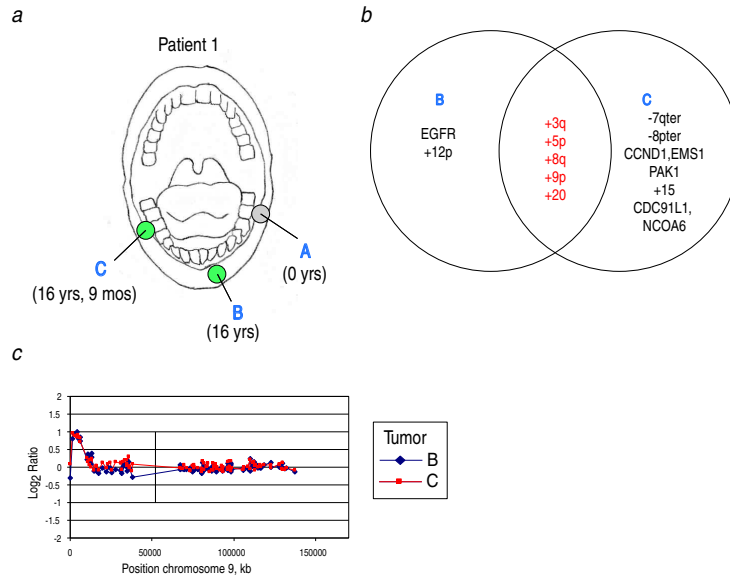


**Figure 7.2:** Array CGH analysis of oral SCCs. Genome wide DNA copy number profile of SCC D from patient 3. Array CGH was performed on arrays comprised of 2464 BAC clones printed in triplicate (HumArray 2.0). Plotted is the mean  $\log_2$  ratio of each triplicate spot for all clones normalized such that the median  $\log_2$  ratio equals 0. Chromosome boundaries are indicated with vertical bars and within each chromosome, clones are ordered based on their genome position according to the July 2003 freeze of the human genome. The use of array CGH allows different types of aberrations to be defined including low-level DNA copy number aberrations of parts of a chromosome, whole chromosome gains or losses and focal high-level amplifications. Chromosomes 5p, 7p and 8q show examples of low-level gains, losses can be observed on chromosomes 3p, 7p, and distal 11q. Whole chromosome gains and losses are seen on chromosomes 14 and 21, respectively. *CCND1*, *EMSI1*, *PAK1*, *RAB30* and *DLG2* are included in several separate high-level amplifications on chromosome 11q13 (Figure 7.5 c).

margins were negative for carcinoma. She remained carcinoma free for 16 years. She then developed pain and a rapidly expanding mass involving the left anterior mandibular gingiva. Biopsy demonstrated an SCC (Figure 7.3, tumor B). She was treated with a resection of the gingiva, mandible and skin of the lower lip. All margins were free of carcinoma. Nine months later she developed an SCC involving the right posterior mandibular gingiva 3 cm posterior to the previous resection margin. (Figure 7.3, tumor C). Gingiva and bone were resected and all margins were free of tumor. She was treated with postoperative radiation therapy. She is now two months following the completion of radiation therapy and there is no evidence of recurrence or a new primary SCC.

## 7.2.2 Patient 2

The patient is a 61 year-old woman, with no history of alcohol or tobacco use, who presented with a painful anterior maxillary gingival lesion (Figure 7.4, tumor B), and a right maxillary palatal lesion (Figure 7.4, tumor A). Biopsies of both lesions demonstrated SCC. The patient underwent a partial maxillectomy. Margins were negative for carcinoma. The patient was then treated with radiation therapy. Sixteen months later she developed an ulcerative and painful lesion around the right mandibular first molar (Figure 7.4, tumor C). Biopsy demonstrated SCC. The patient underwent a resection of the gingiva and a partial mandibulectomy. Margins were negative for carcinoma. Seven years following the maxillary resection she developed a new left maxillary SCC (Figure 7.4, tumor D). A further maxillary resection was performed. All margins were negative. Sixteen months later she developed a right mandibular lesion anterior to the previous site of mandibular resection (Figure 7.4, tumor E). Biopsy demonstrated carcinoma in situ. She underwent resection of the involved gingiva and bone. Fifteen months later an SCC developed in the remaining right maxilla and soft palate (Figure

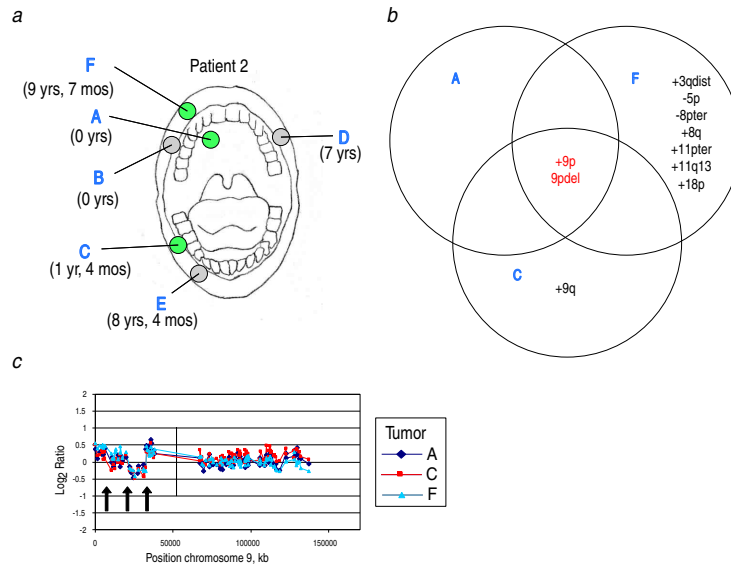


**Figure 7.3:** Copy number aberrations in tumors from patient 1. **a.** Diagram showing locations of multiple lesions and time of occurrence relative to the initial lesion. Tumors analyzed by array CGH are indicated in green. **b.** Venn diagram showing aberrations specific to tumors B and C in black and aberrations shared between these two tumors in red. We were unable to obtain satisfactory array CGH data from tumor A. Amplifications are indicated in bold. The shared aberrations involving gains of 3q, 5p, 8q and 20, which occur frequently in oral SCC could be attributed to a common origin of the tumors or the aberrations could have been selected during tumor development, whereas amplification on distal 9p is unique to this patient. **c.** DNA copy number profile of chromosome 9 in tumor B (blue) and C (red). Copy number ratios on individual clones are plotted from 9pter to 9qter on the July 2003 freeze of the human genome sequence. Both tumors show an amplification of the same amplitude of the distal part of chromosome 9p, which in both tumors is flanked distally by clone CTB-41L13 (9p telomere) and proximally by RP11-50C21. We observed a similar amplification of 9p in one of the 89 randomly selected oral SCCs (specimen 5864). However, while the distal boundary clone of the amplicon, RP11-50C21 was shared, the amplified region extended further in 5864 to include the 9p telomere clone, CTB-41L13. Thus, a 9p amplicon bounded by RP11-50C21 and CTB-41L13 has only been observed in the two tumors from patient 1, providing strong evidence that these two tumors are genetically related.

7.4, tumor F). She underwent a further maxillary resection. While recovering in the hospital she was found to have a pulmonary metastasis and died of her disease

### 7.2.3 Patient 3

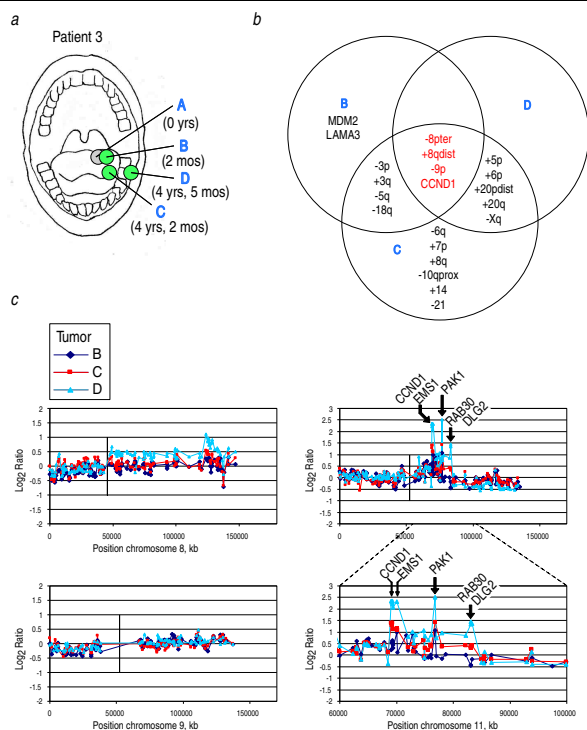
The patient is a 63 year-old woman, with no history of tobacco or alcohol use, who presented complaining of a painful, left tongue ulcer (Figure 7.5, tumor A). Biopsy demonstrated SCC and she underwent a partial glossectomy. Two months later an SCC developed at the site



**Figure 7.4:** Copy number aberrations in tumors from patient 2. **a.** Diagram showing locations of multiple lesions and time of occurrence relative to the initial lesion. Tumors analyzed by array CGH are indicated in green. **b.** Venn diagram showing that the only common aberration between these three tumors is a specific DNA copy number pattern on chromosome 9p, which included both increases and decreases in copy number of parts of 9p (red). Aberrations unique to tumors C and F are shown in black. **c.** DNA copy number profile of chromosome 9 in tumors B (dark blue), C (red) and F (light blue). Copy number ratios on individual clones are plotted from 9pter to 9qter on the July 2003 freeze of the human genome sequence. In all three tumors, both the most proximal and distal parts of 9p show an increase in copy number, while a region in between shows a decrease in copy number. All three tumors exhibit transitions from one copy number level to another at the same locations (arrows) and include two levels of gain on distal 9p as well as the copy number transitions bounding the deletion. In the randomly selected set of 89 oral SCCs, we observed a similar copy number profile with the same deletion breakpoints on 9p in one tumor (specimen 6200). However, no regions of gain were present on 9p in this tumor. Thus, since this particular "signature" copy number change was never observed in 89 randomly selected oral SCCs, it is highly unlikely that this genomic aberration occurred randomly in all three tumors from patient 2.

(Figure 7.5, tumor B). Further resection was performed and all margins were negative. Four years later she developed a painful left posterior tongue lesion that demonstrated SCC (Figure 7.5, tumor C). A further tongue resection was performed. One margin was positive for carcinoma; however a further resection showed no carcinoma. Three months later she developed a left posterior mandibular gingival SCC 3 cm remote from the tongue lesion (Figure 7.5, tumor D). She underwent a resection of the left mandibular gingiva and mandible and was treated with postoperative radiation therapy. She is now seven months following surgery with no clinical evidence of carcinoma.





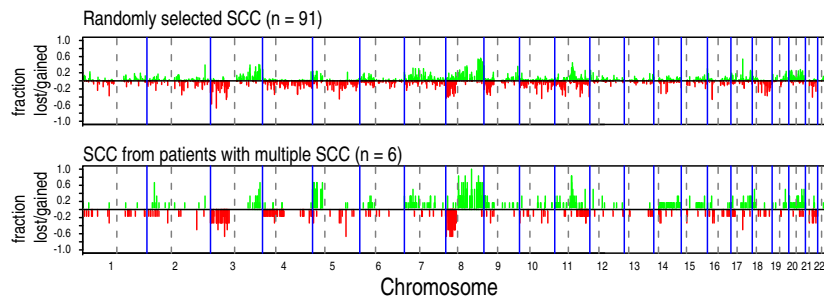
**Figure 7.5:** Copy number aberrations in tumors from patient 3. **a.** Diagram showing locations of multiple lesions and time of occurrence relative to the initial lesion. Tumors analyzed by array CGH are indicated in green. **b.** All three tumors share a loss of 8pter, a gain of the distal part of chromosome 8q, loss of 9p and an amplification of the *CCND1* locus (red). Aberrations unique to each tumor are shown in black and amplifications are indicated in bold. **c.** DNA copy number profiles for chromosomes 8, 9, 11 and an expanded view of the copy number profiles in the *CCND1* region. The whole chromosome copy number profiles show ratios on individual clones plotted from the tip of the p arm to the most distal clone on the q arm on the July 2003 freeze of the human genome sequence. For the expanded view of the *CCND1* region, data are plotted from position 60,000 to 100,000 kb on 11q (July 2003 freeze). In all tumors discrete copy number maxima were observed at *CCND1* and/or *EMSI* and at *PAK1*. However differences included an additional peak centered over *RAB30* and *DLG2* that was present only in tumor D, and a shift in the position of the amplicon maximum from *EMSI* in tumor B to center on *CCND1* in tumor C and include both *CCND1* and *EMSI* in tumor D. In addition there is an apparent increase in amplitude in later tumors.

## 7.3 Results

### 7.3.1 Array CGH analysis of tumors from patients with multiple SCC

Since the majority of patients do not develop second primaries, we first determined whether or not tumors in patients presenting with multiple second primaries represented a distinct class of oral SCC. We compared the copy number aberrations occurring in a single tumor

from each of six patients with a history of second primary tumors with those observed in a set of 89 oral SCCs selected randomly from the UCSF Oral Pathology Tissue Bank (Snijders *et al.*, in preparation). In both sets of tumors, we found similar losses and gains of genomic regions, including frequent losses of 3p, 4, 5q, 8p, 9p, 10, distal 11q, 18q and 21 and gains of 3q, 5p, 6p, 7p, 8q, proximal 11q, 19q and 20 (Figure 7.6). In both tumor sets amplifications of *EGFR* and *CCND1* were present as were other amplifications. Thus, we found no evidence of genomic differences in tumors arising in patients with multiple oral SCCs compared to other oral SCCs.



**Figure 7.6:** Frequency of gains (indicated in green, ranging from 0 to 1) and losses (indicated in red, ranging from 0 to -1) for each clone for 89 randomly selected SCCs (top) and six SCCs from each of six patients with a history of second primary tumors (bottom).

### 7.3.2 Copy number alterations present in multiple tumors from individual patients

For all patients we observed copy number alterations that were shared amongst all the analyzed tumors from an individual patient (Figures 7.3, 7.4, 7.5), as well as other copy number changes that were either unique to one of the tumors or shared by only two of the three analyzed tumors (patient 3). Some of the shared aberrations were seen frequently in oral SCCs (e.g. gain of 3q, 5p or amplification of *EGFR* or *CCND1*), others formed a unique signature of the multiple tumors from an individual, because they were never observed in the set of 89 randomly selected tumors. For example, in patients 1 and 2, the signature aberrations were unique copy number changes on chromosome 9p, amplification of distal 9p (Figure 7.3, patient 1) and both gains of regions on 9p and a deletion encompassing *CDKN2A* in the middle of 9p (Figure 7.4, patient 2). These signature copy number changes provide the clearest evidence that the multiple tumors from one individual are genetically related.

Shared aberrations in the three tumors from patient 3 included losses of 8pter and 9p, gain of distal 8q and amplification of *CCND1*, *EMSI* and *PAK1* at 11q13 (Figure 7.5). Other aberrations were shared between only two of the three tumors (B and C or C and D) and some aberrations were present in only one tumor. While each of the shared aberrations in the tumors from patient 3 were all observed frequently in oral SCCs, none of the 89 randomly selected tumors shared all of these aberrations. For example, amplification of *CCND1* was present in 10/89 randomly selected tumors. Deletion of some portion of 9p was observed in five of these, but none of the five also had a gain of distal 8q. Furthermore, the detailed

structure of the copy number profiles and the positions of the copy number transitions defining the copy number losses and gains on chromosomes 8 and 9 never co-occurred in any of the 89 randomly selected oral SCCs. On the other hand, some variation in the detailed structure of the copy number profile across the amplified region on chromosome 11q13 is apparent (Figure 7.5 c), which is likely a reflection of greater instability of chromosomes carrying amplified DNA as observed in model experiments in cells in culture (Snijders et al., 2003 and unpublished observations).

## 7.4 Discussion

We have demonstrated genetic similarities amongst the multiple SCCs from each patient, which we conclude is evidence that they represent clonally related second primary tumors. Other explanations for sharing of aberrations include selection because the aberrations were particularly advantageous for oral SCC development in general or because an individual's genetic susceptibility favored these particular changes. However, these alternatives are unlikely for two reasons. First, the signature copy number aberrations were not frequent events in the set of 89 randomly selected oral SCCs. Second, the detailed structure of the copy number profiles of the shared aberrations was conserved, which would have required that the chromosomal breakage and rearrangement events that result in copy number alterations had occurred repeatedly at the same places in the genome (i.e. within ~1-2 Mb).

A history of multiple genetically related SCCs in these patients points to either the presence of abnormally expanded fields of premalignant cells (Figure 7.1 b) or a mechanism involving migration of malignant or premalignant epithelium (Figure 7.1 c) and raises the question as to how such large fields develop. Progenitor cells in stratified squamous epithelia, including the oral cavity are situated in the basal layer (Squier and Kremer, 2001; Owens and Watt, 2003) and in skin, form clonal proliferating units measuring ~ 2 mm in diameter (Chaturvedi et al., 2002). However, studies of the extent of fields of abnormal cells in patients with head and neck tumors or the size of patches of *TP53* mutant cells in normal appearing sun exposed skin indicate that the affected area can be extensive (Wolf et al., 2004; Waridel et al., 1997; Jonason et al., 1996; Califano et al., 2000). Observations on UVB induced mutant p53 patches in mouse skin indicate that the rate limiting step for clonal expansion is acquisition of the capability to escape the boundaries of the stem cell unit. Furthermore, expansion of mutant patches is not cell autonomous, but requires continuous UVB exposure (Zhang et al., 2001b). Thus, in the patients studied here, currently unknown environmental factors might be providing a continual stimulus for expansion of mutant cells in the oral cavity in a manner similar to UVB in skin.

Currently clinical management of oral SCC patients follows an established protocol, whereby patients are treated with surgery and possibly radiation therapy, then followed clinically to identify a second primary SCC early. However, there appears to be a distinct class of patients at high risk for multiple primary SCC development. These patients are likely to benefit from application of molecular analysis techniques to monitor clinically and histologically normal mucosa for fields of abnormal cells. Thus, identification of alterations characteristic of a primary SCC using array CGH followed by performing fluorescent *in situ* hybridization (FISH) on surgical or brush biopsies would allow treatment to be targeted based on molecular field analysis and would be a step forward in the management, control and prevention of oral SCC.

## 7.5 Methods

### 7.5.1 Isolation of genomic DNA

Tumor rich regions were dissected from 15 consecutive 10  $\mu\text{m}$  formalin fixed paraffin embedded tissue sections from routine surgical biopsies. Sections were deparaffinized by incubating for 10 min. on a rocking table in three changes of 1 mL xylene. After centrifugation and removal of the supernatant, 1 mL of absolute ethanol was added and sections were incubated for 15 min. at room temperature. Sections were air-dried after centrifugation and removal of ethanol, followed by overnight incubation in 500  $\mu\text{L}$  digestion buffer (100 mM NaCl, 25 mM EDTA, 10 mM Tris-HCl, 0.5% SDS in  $\text{H}_2\text{O}$ ). Sections were digested by adding 400-500  $\mu\text{g}$  of proteinase K per day for three consecutive days and incubation at 55°C. DNA was extracted using phenol-chloroform-isoamylalcohol (25:24:1) and ethanol precipitated with ammonium acetate in the presence of 100  $\mu\text{g}$  glycogen. The precipitate was collected by centrifugation and the pellet was air-dried and dissolved in 15  $\mu\text{L}$   $\text{H}_2\text{O}$ .

### 7.5.2 Array CGH

Array CGH was carried out as described previously (Snijders *et al.*, 2003). Briefly, genomic DNA (600 ng) was labeled by random priming to incorporate Cy3- or Cy5 dCTP in a 50  $\mu\text{L}$  reaction. Labeled test (600 ng) and reference DNAs (300 ng) together with 100  $\mu\text{g}$  human Cot-1 DNA were hybridized for ~48 hrs at 37°C to arrays of 2464 BAC clones each printed in triplicate (HumArray2.0, UCSF Comprehensive Cancer Center Microarray Core). A custom built CCD camera system was used to acquire 16 bit 1024x1024 pixel DAPI, Cy3 and Cy5 images (Pinkel *et al.*, 1998). Image and data analysis were carried out using UCSF SPOT (Jain *et al.*, 2001) and SPROC software, as described previously (Snijders *et al.*, 2003).

## 7.6 Acknowledgements

This work was supported by NIH grants CA90421, CA94407, CA95231, DE13904 and Tobacco-Related Disease Research Program grant 11RT-0141. BLS is an appointee of the Western Oral Research Consortium (NIH K12 DE14609).

## Chapter 8

# Rare Amplicons Implicate Frequent Misspecification of Cell Fate in Oral Squamous Cell Carcinoma

Antoine M. Snijders<sup>1\*</sup>, Brian L. Schmidt<sup>2\*</sup>, Jane Fridlyand<sup>3</sup>, Nusi Dekker<sup>4</sup>, Daniel Pinkel<sup>5</sup>,  
Richard C. K. Jordan<sup>4,6</sup> and Donna G. Albertson<sup>1,5</sup>

\*These authors contributed equally.

<sup>1</sup>Cancer Research Institute, University of California San Francisco, San Francisco CA, USA

<sup>2</sup>Department of Oral and Maxillofacial Surgery, University of California San Francisco,  
San Francisco CA, USA

<sup>3</sup>Department of Epidemiology and Biostatistics, University of California San Francisco,  
San Francisco CA, USA

<sup>4</sup>Department of Stomatology, University of California San Francisco, San Francisco CA,  
USA

<sup>5</sup>Department of Laboratory Medicine, University of California San Francisco, San Francisco  
CA, USA

<sup>6</sup>Department of Pathology, University of California San Francisco, San Francisco CA, USA

Submitted

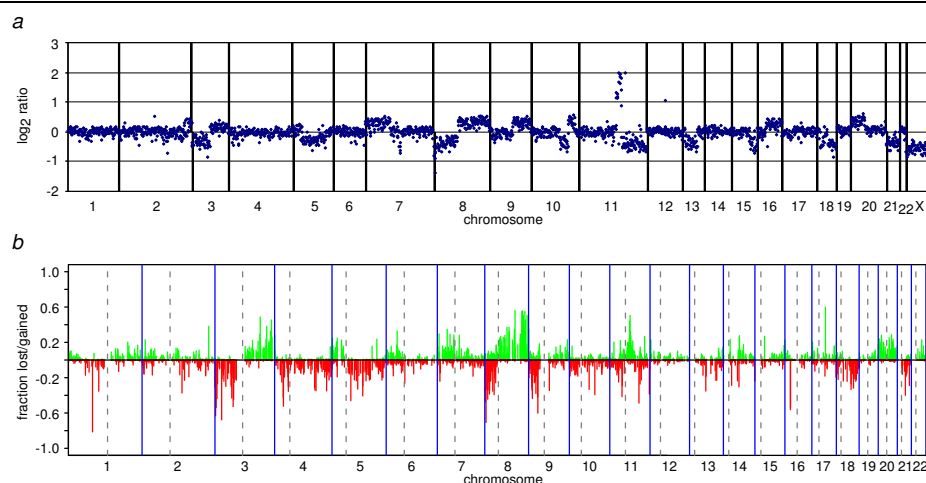
## 8.1 Abstract

Genomes of solid tumors are characterized by gains and losses of regions, which may contribute to tumorigenesis by altering gene expression. Often the aberrations are extensive, encompassing whole chromosome arms, which makes identification of candidate genes in these regions difficult. Here, we focused on narrow regions of gene amplification to facilitate identification of genetic pathways important in oral squamous cell carcinoma (SCC) development. We used array comparative genomic hybridization (array CGH) to define minimum common amplified regions and then used expression analysis to identify candidate driver genes in amplicons that spanned < 3 Mb. We found genes involved in integrin signaling (*TLN1*), survival (*YAP1*, *BIRC2*), and adhesion and migration (*TLN1*, *LAMA3*, *MMP7*), as well as members of the Hedgehog (*GLI2*) and Notch (*JAG1*, *RBPSUH*, *FJX1*) pathways to be amplified and over-expressed. Deregulation of additional members of the hedgehog and notch pathways (*HHIP*, *SMO*, *DLL1*, *NOTCH4*) implicates cell fate misspecification in oral SCC development.

## 8.2 Results and Discussion

Although controversy surrounds the functional significance of the many DNA copy number changes seen in tumor genomes, amplifications, defined as regions of focal high level copy number change (Snijders *et al.*, 2001), are likely to represent alterations continuously under selection for tumor growth, since studies indicate that the amplified DNA is unstable. On the other hand, there are many ways to up-regulate gene expression in addition to increasing the copy number of the gene (Albertson *et al.*, 2003). Bearing in mind these observations, we hypothesized that even rarely occurring amplicons would be informative regarding genes important in tumor development. To apply this logic to oral squamous cell carcinoma (SCC), we used array comparative genomic hybridization (array CGH) to obtain genome-wide information on copy number alterations (Figure 8.1) in 89 SCC taken from four different sites in the oral cavity (Table 8.1, Web Table A). This analysis revealed a number of frequent low level gains and losses (Figure 8.1b) and 18 regions of recurrent amplification (Table 8.2), a number of which contained genes found previously to be amplified and/or overexpressed in oral SCC. Hierarchical clustering of the array CGH data of all 89 tumors revealed three main clusters (Appendix C.1, Web Figure A), one of which is significantly enriched with tumors with mutations in *TP53* exons 5-8 (Fisher exact test p-value = 0.001). Low level copy number alterations significantly associated with *TP53* mutational status after maxT (Westfall and Young, 2003) correction for multiple testing included -8p, +distal 8q, -10q, -11q and -18q (Appendix C.2, Web Figure B). In addition, we observed that mutation of *TP53* was positively correlated with amplification of *CCND1* (Fisher exact test p-value = 0.009), confirming a previous report (Mineta *et al.*, 1997) and with amplification of *EGFR* (Fisher exact test p-value = 0.036).

We then focused on the nine amplicons with boundaries spanning less than 3 Mb (Figure 8.2). We identified two regions containing genes frequently amplified and/or over-expressed in oral SCC, including *EGFR* (n=10 cases) at 7p11.2, and/or two separate, but co-amplified regions at 11q13 encompassing *CCND1*, *FGFR3*, *FGFR4* and *EMS1* in a 1.5 Mb amplicon and *PAK1* in a separate 0.9 Mb amplicon (n=10 cases). The remaining seven amplicons did not contain well-established oncogenes in oral SCC. Therefore, to identify the driver genes in these amplicons, we evaluated transcript levels of candidate genes mapping in the minimum



**Figure 8.1:** Genome-wide analysis of copy number aberrations in 89 oral SCC. **(a)** Genome-wide DNA copy number profile of one oral SCC. Plotted is the normalized  $\log_2$  ratio for each clone sorted by chromosome and ordered according to genome position from the p-arm to the q-arm. Multiple low-level DNA copy number aberrations are present, including loss of chromosome 3p, 5q, 8p, 18 and 21 and gain of chromosome 7q, 8q, 9q and 20p. This tumor also has an amplification at 11q13 including *CCND1*. **(b)** Frequency of gains, indicated by the green bars ranging from 0 to 1, and losses, indicated by the red bars ranging from 0 to -1, in 89 oral SCCs for each individual clone. The most recurrent regions of DNA copy number loss are on chromosomes 3p, 4, 5q, 8p, 9p, 18 and 21, while recurrent regions of copy number gains are on chromosomes 3q, 8q, 11q and 20 in general agreement with previous studies of oral SCC (Wolff *et al.*, 1998; Huang *et al.*, 2002).

Site (n)	Gender		Mean Age (range)
	Male	Female	
Buccal Mucosa (17)	7	10	64 (35-90)
Floor of Mouth (17)	10	7	68 (49-91)
Gingiva (21)	11	10	70 (35-88)
Tongue (34)	14	20	52 (26-78)

**Table 8.1:** Patient characteristics

amplicon in a set of 24 tumors using the quantitative reverse transcription polymerase chain reaction (RT-PCR) and compared the expression levels to normal samples from the tongue (n=2) and buccal mucosa (n=1). These analyses, summarized in Figure 8.3, indicate greater than two-fold up-regulation of expression compared to normal tissue of *GLI2* at 2q14.2 (11/23 tumors), *RBPSUH* at 4p15.2 (4/12 tumors, respectively), *TLN1* at 9p13.3 (4/22 tumors), *FJX1* at 11p13 (15/15 tumors), *YAP1*, *BIRC2*, and *MMP7* at 11q22 (4/21, 4/20, 17/20 tumors, respectively), *LAMA3* at 18q11.2 (20/23 tumors) and *JAG1* at 20p12.2 (10/23 tumors). In

all cases, we observed increased expression of these genes when they were amplified, but more frequently they were over-expressed without amplification, suggesting that they are likely candidate driver genes for amplification of their respective amplicons. We consider *BIRC3* and *KIAA0746* less likely to be driver genes, since we observed little up-regulation of these genes when amplified. Furthermore, they are expressed in lymphatic tissue and thus, over-expression in tumor samples probably reflects inflammatory infiltrate (Web Table B). On the other hand, we observed high levels of expression of *FJX1* in lymph nodes, but not in a buccal mucosa sample with lymphocytic infiltrate. Since *FJX1* was highly up-regulated when amplified, we retained *FJX1* as a candidate driver gene for amplification at 11p13.

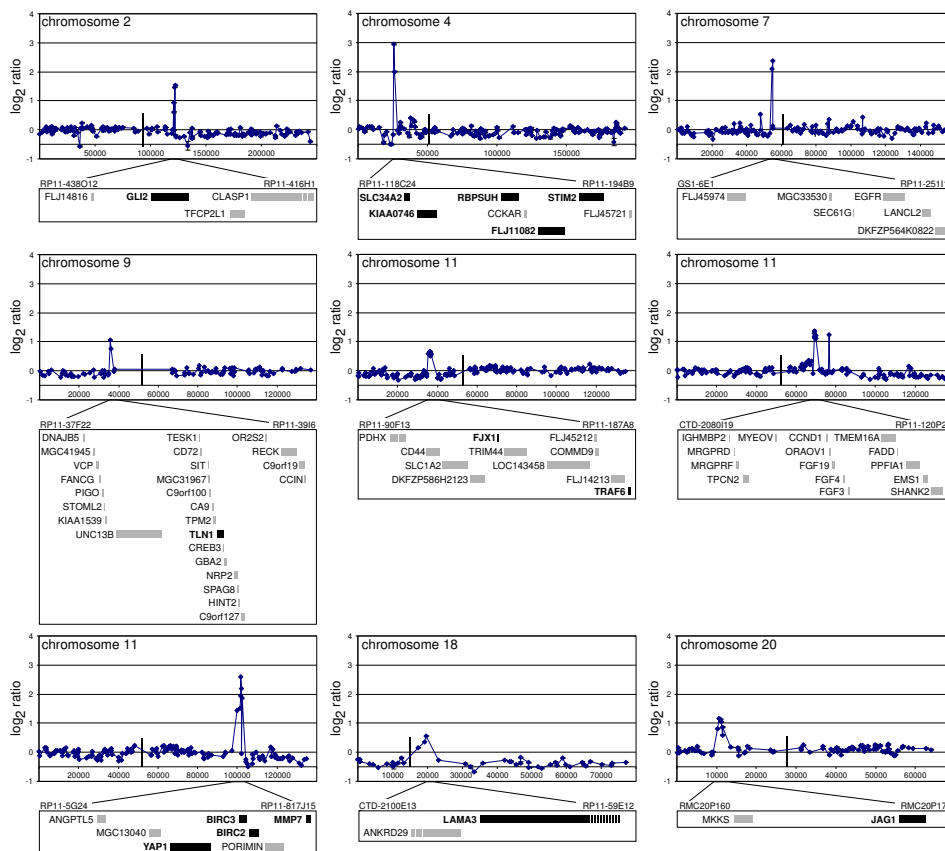
Tumors	Chrom.	Amplicon Copy Nuber Range (log <sub>2</sub> ratio)	Proximal flanking clone	STS	Start (bp)	Distal flanking clone	STS	End (bp)	Size (Mb)	Candidate Genes
4756, 6929	2q14.2	0.5 - 1.5	RP11-438O12		121,225,342	RP11-416H1	D2S343	122,456,621	1.2	GLI2
3103, 3482	3q24-25	1.0 - 1.8	RP11-172E23	D3S1557	146,527,612	RP11-65L11	AFM277W99	152,627,100	6.1	TMSF1
3883, 4052, 4961	4p15.2	0.7 - 2.9	RP11-118C24	SHGC-24618	25,331,058	RP11-194B9	SHGC4-273	28,103,518	2.8	RBPSUH, STM2
5918, 6382	5p13.2	0.7 - 0.9	RP11-67P13	AFMA297WAS	32,233,715	RP11-9G14	D5S634	41,329,416	9.1	RAD1, SKP2, IL7R
1424, 5730	6q12	1.0	RP11-277K21	AFMB291ZB5	64,091,403	RP11-2M9	AFM295TB3	71,712,187	7.6	PTPA41, EGF11
1067, 2042, 3816, 4397, 5463, 5699, 5793, 5832, 5877, 6799, 6988, 3800, 5790	7p11.2	0.7 - 2.6	GS1-4E1		53,465,093	RP11-251115	D7S499	55,349,162	1.9	EGFR
3144, 5918	7q21.2	2.3 - 2.6	CTB-141D22		90,029,465	CTD-2007G21	U31384	93,098,318	3.1	FZD1, CDK6
3482, 3501, 5864, SCC094	8p12	1.1 - 1.2	RP11-210F15	SHGC-20486	36,452,678	RP11-262I23	SHGC-12674	39,744,917	3.3	BAG4, FGFRI1, TACC1, ADAM9
1067, 6362, 6818	9p24.1	0.9 - 1.4	RP11-12N24	SHGC-34067	5,223,942	RP11-50C21	SHGC-18065	10,431,717	5.2	UHRF2, JMJ2C, PTPRD
280, 4833, 9420, 6362, 6831	9p13.3	0.6 - 2.1	RP11-37F22	SHGC-35988	34,971,713	RP11-39I6	SHGC-32888	36,174,984	1.2	TLN1
280, 2042, 2962	11p13	0.6 - 1.1	RP11-90F13	AFMA081T05	34,957,558	RP11-187A8	SHGC-6028	36,559,560	1.6	CD44, FJX1, TRAF6
2042, 4833, 5730, 5771, 4397, 5790, 6508, 6672, 6799, 6988, 6672, 6988	11p11.2	0.6 - 1.0	RP11-102E22	SHGC-13806	44,278,905	CTD-2244P3		47,245,301	3.0	MAPK8IP1, BHC80
	11q13.3	0.7 - 3.6	CTD-208019	RH7839	68,482,959	RP11-120P20	SHGC-4518	70,129,383	1.6	CCND1, FGF3, 4, 19, EMS1
	11q13.5	1.4 - 1.8	CTC-352E23	RH52308	76,097,938	RP11-98G24	SHGC-31540	77,013,406	0.9	PAK1
1067, 2861, 3800, 5833, 6420	11q22	0.7 - 2.6	RP11-5G24	SHGC-10856	101,181,950	RP11-817J15	SHGC-11011	101,955,349	0.8	YAP1, BIRC2, BIRC3, MMP7
3800, 4961, 5463, 5814, 6362	12q15	0.4 - 1.3	RP11-5J6	SHGC-3797	66,882,015	RP11-92P22	SHGC-35465	74,052,886	7.2	MDM2, PTPRR
2861, 2868, 6508, 6988	18q11.2	0.4 - 1.2	CTD-2100E13	STSG21909	19,275,988	RP11-59E12	AFM164ZC1	19,809,522	0.5	LAMA3
1067, 3482, 4961	2012.2	0.5 - 1.1	RMC20P160	WI-7829	10,282,059	RMC20P178	D20S186	11,518,795	1.2	JAG1

**Table 8.2:** Recurrent amplicons in 89 oral SCC

Based on the known functions of the candidate genes, we can propose plausible roles in tumorigenesis, including deregulation of transduction of integrin signaling (*TLN1*) (Nayal *et al.*, 2004), opposition to apoptosis (*YAP1* and *BIRC2*), and adhesion and migration (*TLN1*, *LAMA3*, *MMP7*) (Nayal *et al.*, 2004; Lohi, 2001). However, we were particularly interested in the observed amplification and overexpression of *GLI2*, *JAG1*, *RBPSUH* and *FJX1*, members of the Hedgehog and Notch pathways. These ontogenetic networks function in cell fate specification and are widely conserved across phyla. Deregulated Hedgehog signaling is a hallmark of basal cell carcinomas of the skin (Ruiz i Altaba *et al.*, 2002) and odontogenic keratocysts in the oral cavity (Ohki *et al.*, 2004), both of which are phenotypically distinct from SCC. On the other hand, Notch signaling promotes differentiation in skin (Lefort and Dotto, 2004) and loss of function of Notch is permissive for tumor formation in mouse models (Nicolas *et al.*, 2003). Although one study found up-regulation of Wnt and Notch signaling pathways in head and neck SCC (Leethanakul *et al.*, 2000), the Hedgehog and Notch pathways have not been previously characterized in oral squamous cell tumorigenesis. Therefore, we investigated whether expression levels of other members of the hedgehog and notch pathways were altered in oral SCC.

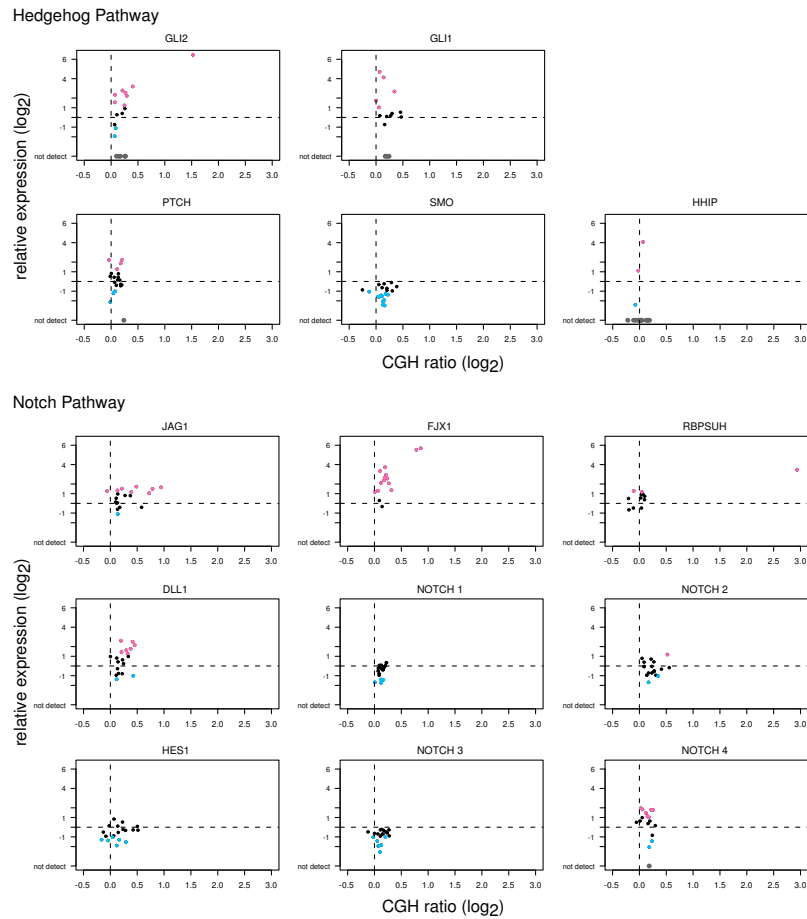
Hedgehog signaling involves secretion of the hedgehog morphogens, Sonic (*SHH*), Indian (*IHH*) and Desert (*DHH*) hedgehog, which interact with Patch to relieve inhibition of Smoothed, resulting in activation of the transcription factors, *GLI1*, *GLI2* and *GLI3* (Nicolas *et al.*, 2003). We observed that *GLI2* expression was positively correlated with expression of *GLI1* ( $R^2=0.7$ ) and *PTCH* ( $R^2=0.5$ ), consistent with observations in other systems in which it has been shown that *GLI2* transcriptionally regulates expression of these genes. We observed that *SMO* was more than two-fold down regulated in the majority of tumors (11/19). Limited





**Figure 8.2:** Overview of recurring DNA copy number amplifications in 89 oral SCC determined by genome-wide array CGH. We compared the extent of amplicons in a series of tumors to determine the minimum recurrent region of amplification to facilitate identification of potential driver genes underlying each aberration. We show typical copy number profiles for nine recurrent regions of amplification that did not exceed 3 Mb (range 0.5 - 2.8 Mb) observed on chromosomes 2, 4, 7, 9, 11 (3 regions) and 18. For all chromosomes, we plotted the data according to position on the chromosome in kilobases using the July 2003 freeze of the draft human genome sequence (<http://www.genome.ucsc.edu/>) and indicate the position of the centromere by a vertical bar. The expanded view of the minimum common region of amplification, shown underneath each amplicon includes the most proximal and distal flanking clones and the candidate genes present in the region (UCSC July 2003 freeze). Since driver gene(s) are likely to be over-expressed at the transcript level, we performed quantitative RT-PCR analysis of candidate driver genes in seven of the nine recurrent regions of amplification in a subset of tumors including amplifying tumors. We highlighted assayed genes in bold and genes that we did not assay in gray.

analysis of oral SCC cell lines has shown that *SHH* is expressed in several and further that the pathway is likely to be active in at least one cell line, since growth can be inhibited by exposure to the Smoothened inhibitor cyclopamine (Nishimaki *et al.*, 2004). However, we



**Figure 8.3:** Summary of copy number and expression of amplified genes and associated pathway members. For each gene we show expression levels in a series of tumors relative to the average expression in several normal oral tissue samples (Web Table B) versus DNA copy number for each corresponding gene observed by array CGH on a BAC clone closest to or encompasses the gene. We highlighted genes which were more than two-fold over-expressed in pink, genes downregulated more than two-fold in blue and genes whose expression was not detected in gray. Genes which expression level did not differ from normal oral epithelial samples are indicated in black. We observed that amplified genes (*GLI2*, *JAG1*, *FJX1*, *RBPSUH*) were always over-expressed, but more frequently genes were over-expressed without amplification (*GLI2*, *GLI1*, *JAG1*, *FJX1*, *RBPSUH*, *DLL1*, *NOTCH4*).

found no expression of *SHH*, *IHH* or *DHH* in normal tissue and only rarely detected expression in tumors (0/19, 0/20 and 1/5 tumors, respectively). In addition, although we detected expression of *HHIP*, a negative regulator of hedgehog signaling in normal tissue, we found no expression in the majority of tumors (14/18). Taken together these observations suggest that endogenous ligand activation of Hedgehog signaling is not inducing *GLI2* up-regulation in oral SCC, in contrast to tumors from tissues of endodermal origin, including esophagus,

stomach and pancreas (Watkins and Peacock, 2004). On the other hand, loss of function mutations of *Hip1* in mice result in up-regulation of hedgehog signaling (Chuang *et al.*, 2003), suggesting that down regulation of *HHIP* may contribute to deregulation of hedgehog signaling in oral SCCs.

Notch signaling is initiated by interaction of ligands (e.g. *DLL1*, *JAG1*) with membrane-bound Notch resulting in cleavage of the Notch protein. The released notch intracellular domain enters the nucleus and interacts with the transcription factor, *RBPSUH* leading to expression of "hairy enhancer of split" genes (Lefort and Dotto, 2004; Weng and Aster, 2004). Although Notch signaling promotes differentiation in skin (Lefort and Dotto, 2004), in oral SCC, we found amplification and over-expression of *JAG1* and *RBPSUH*, two genes central to notch signaling. In addition, we observed amplification of *FJX1*, the human homologue of four-jointed, which acts in *Drosophila* as a second signal downstream of Notch in leg development (Buckles *et al.*, 2001) and in regulation of polarity in ommatidial development (Zeidler *et al.*, 1999). The function of *FJX1* in humans is currently unknown, but amplification of this gene together with *JAG1* and *RBPSUH* in oral SCC suggests that deregulation of Notch signaling is likely contributing to development of these tumors. Therefore we examined the expression of other Notch pathway genes in tumors and found that expression of *HES1* and *NOTCH1* increased with increasing expression of *JAG1* and *DLL1* (Web Table B), suggesting that Notch signaling was active in the tumors. We observed that *DLL1* and *NOTCH4* were up-regulated two-fold relative to normal tissue (8/21 and 8/19 tumors, respectively), whereas we found a tendency for *NOTCH1*, 2 and 3 and *HES1* to be expressed at lower levels. Up-regulation of *NOTCH4* is associated with breast and mammary cancer (Weng and Aster, 2004). We also observed up-regulation of *JAG1* expression in moderate dysplasias (3/7), but not in mild dysplasias (0/4). In contrast, expression of the inhibitor of Notch signaling, Manic Fringe was up-regulated more than two-fold relative to normal tissue in all dysplasias (2-9 fold), but was less frequently over-expressed in tumors (2-5 fold in 9/13 tumors). These observations suggest that Notch signaling is activated in the later stages of oral cancer development. Similar up-regulation of Notch activity is associated with the transition from cervical intraepithelial lesions to invasive carcinoma (Veeraraghavalu *et al.*, 2004; Gray *et al.*, 1999). Thus, it appears that activated Notch signaling contributes to development of squamous cell carcinoma in various keratinizing epithelia. Skin and oral epithelia are continually renewed and show a developmental gradient from the basal layer to the outermost layer where cells are shed. In the oral cavity, epithelial stem cells located in the deep rete ridges divide rarely to give rise to transit amplifying cells, which are committed to differentiation, but divide a limited number of times prior to withdrawing from the cell cycle (Hume and Potten, 1979). Oral SCC progenitor cells are likely to be stem cells that have acquired appropriate mutations to allow them to proliferate abnormally and/or to render transit amplifying cells resistant to differentiation. Amplification and/or over-expression of members of the hedgehog and notch pathways in tumors suggest that deregulation of these pathways plays a role in misspecification of oral epithelial cell fates leading to tumor development. Since a number of the involved genes may be expressed in various forms with different growth promoting or repressing functions (Aho, 2004; Ascano *et al.*, 2003) and the outcome of notch signaling is context dependent (Weng and Aster, 2004), identification of the precise roles of these pathways in oral cancer awaits further characterization of the over-expressed proteins and their pathways. Oral SCC remains a significant health problem (Parkin *et al.*, 1999). The 5-year survival rate, at 40%, is among the worst of all sites in the body and has not improved over the past 40 years (Parkin *et al.*, 1999). Oral epithelial dysplasia often precedes SCC

development. Hence, there is both a need for more effective therapies and also the opportunity to recognize premalignant lesions and initiate chemoprevention. The identification of active hedgehog and notch signaling in oral SCC suggests the possibility of applying new therapeutic approaches targeting members of these pathway (Watkins and Peacock, 2004).

## 8.3 Materials and Methods

### 8.3.1 Tumor Samples

The oral SCC, dysplasia and normal tissue specimens and associated clinical data were obtained through the UCSF Oral Cancer Tissue Bank. Normal cervical lymph nodes were removed as part of surgical treatment for benign oral disease and were also obtained through the UCSF Oral Cancer Tissue Bank. All tissues were fixed in formalin prior to processing in paraffin. Patient consent was obtained for use of all specimens. (Table 8.1 and Web Table A). Prior to nucleic acid extraction, we stained the first and last sections with hematoxylin and eosin. We examined these sections to confirm the diagnosis of SCC prior to collection of material for nucleic acid extraction and to estimate the normal cell content of the regions selected for dissection, which varied from 60-90% epithelial cells.

### 8.3.2 Isolation of DNA

We dissected tumor rich regions from 15 consecutive 10  $\mu$ M formalin fixed paraffin embedded tissue sections from routine surgical biopsies and tumor resections. We de-paraffinized the sections by using three 10 min. incubations with 1 mL xylene on a rocking table. We removed the xylene following centrifugation of the sections incubated the sections in 1 mL of absolute ethanol for 15 min. at room temperature. After removal of the ethanol, we air-dried the sections and then incubated them over night in 500  $\mu$ L digestion buffer (100 mM NaCl, 25 mM EDTA, 10 mM Tris-HCl, 0.5% SDS in H<sub>2</sub>O). We treated the sections with 400-500  $\mu$ g of Proteinase K each day for three consecutive days at 55°C and then extracted DNA using phenol-chloroform-isoamylalcohol (25:24:1). We precipitated the DNA with ethanol and ammonium acetate in the presence of 100  $\mu$ g glycogen and collected the pellet by centrifugation. We dissolved the air-dried pellet in 15  $\mu$ L H<sub>2</sub>O and determined the DNA concentration by fluorometry.

### 8.3.3 Array CGH

We carried out array CGH as described previously (Snijders *et al.*, 2001). Briefly, to label genomic DNA (600 ng), we used random priming to incorporate Cy3- or Cy5 dCTP in a 50  $\mu$ L reaction. We hybridized labeled test (600 ng) and reference DNAs (300 ng) together with 100  $\mu$ g human Cot-1 DNA for ~48 hrs at 37°C to arrays of 2464 BAC clones each printed in triplicate (HumArray2.0, UCSF Comprehensive Cancer Center Microarray Core) (Snijders *et al.*, 2003). We acquired 16 bit 1024x1024 pixel DAPI, Cy3 and Cy5 images using a custom built CCD camera system and carried out image and data analysis using UCSF SPOT. We used SPROC software to automatically filter the data to reject data points based on low DAPI intensity, low correlation between Cy3 and Cy5 within each segmented spot and low reference/DAPI signal intensity. We declared as missing observations with no replicates or with standard deviation of the replicates greater than 0.2. We screened out clones for the

following reasons; data was missing in more than 15% of the samples, ratios on the clones had shown a median absolute value  $> 0.2$  in the normal samples, clones were not mapped on the genome sequence, or they were known as common copy number polymorphisms. For each tumor, we plotted the data in genome order as the mean  $\log_2$ ratio of the replicate spots for each clone normalized to the genome median  $\log_2$ ratio.

### 8.3.4 Statistical Analysis

We estimated the experimental variability of each profile, *sd*, by taking the median of the median absolute deviations of the measurements on clones with the same copy number in that profile (Fridlyand *et al.*, 2004). We declared a clone gained (lost) if its absolute value exceeded 2.5 times the *sd* for a given profile. We identified high level amplifications by considering the magnitude and the width of the peak (Fridlyand *et al.*, 2004). We used the Fisher exact test to test for independence of *TP53* mutation status with *EGFR* and *CCND1* amplifications and with the cluster assignments. We used t-statistic with pooled variance to test for the differential copy number between *TP53* mutant and wild type samples. when comparing groups of tumors with or without *TP53* mutation by testing each clone for differential copy number using a t-statistic with pooled variance. To assess significance for individual clones, we obtained the maxT-adjusted permutation-based p-values (Westfall and Young, 2003). We declared a clone to have significantly different copy number between the two groups if its adjusted p-value was less than 0.05. Thus, there is less than 5% chance of obtaining one false positive result or more.

### 8.3.5 Quantitative RT-PCR

We isolated RNA from consecutive sections of formalin fixed paraffin embedded tumor resection or biopsy specimens<sup>26</sup> and performed real-time quantitative RT-PCR as previously described<sup>27</sup> in the UCSF Comprehensive Cancer Center Genome Analysis Shared Resource Facility using ABI Assays-on-Demand Expression assays (Web Table B).

### 8.3.6 *TP53* sequencing

We amplified exons 5-8 of *TP53* from genomic DNA of 76 tumors (24 tongue, 16 buccal mucosa, 17 floor of mouth, 19 gingiva) and carried out cycle sequencing as described previously<sup>28</sup> using modified primer sequences for exons 7 (forward: 5'-TGCCACAGGTCTCCC CA-3' and reverse: (5'-ATGGAAGAAATCGGTAAGAGGTG-3')) and 8 (forward: 5'-CCTT ACTGCCTCTTGCTTC-3' and reverse: 5'-CATAACTGCACCCCTTGTC-3').

## 8.4 Acknowledgements

This work was supported by NIH grants CA90421, CA94407, CA95231 and DE13904, and Tobacco-Related Disease Research Program grant 11RT-0141. BLS is an appointee of the Western Oral Research Consortium (NIH K12 DE14609).



## Appendix A

# Supplementary information: Assembly of microarrays for genome-wide measurement of DNA copy number by CGH

### A.1 Methods

#### A.1.1 Specimens

We obtained GM03563, GM00143, GM05296, GM07408, GM01750, GM03134, GM13330, GM03576, GM01535, GM07081, GM02948, GM04435, GM10315, GM13031 and GM01524 cell strains from the NIGMS Human Genetics Cell Repository (Coriell Institute for Medical Research), and cultured them under the recommended conditions. We cultured cell lines BT474, HCT116, T47D, MPE600, COLO320, SW837, MDA-MB-231, MDA-MB-453 and HT29 under the recommended conditions. To isolate DNA from confluent cell cultures, we collected cells from T75 flasks and then incubated them overnight at 55°C in a 3 ml solution containing 0.01 M Tris, pH 7.5, 0.001 M EDTA, pH 8, 0.5% SDS and 0.1  $\mu\text{g}/\mu\text{l}$  proteinase K. We added 1 ml of saturated NaCl solution and 10 ml of ethanol to the cell lysate and collected the DNA by spooling. We removed excess liquid and then dissolved the DNA in 1 ml of H<sub>2</sub>O. We isolated DNA from breast tumor specimens as described previously (Pinkel *et al.*, 1998).

#### A.1.2 Genomic clones

We selected the majority of the clones for the arrays from the set of cytogenetically mapped BACs reported previously (Cheung *et al.*, 2001) and obtained these clones from the Roswell Park Cancer Institute. We obtained additional clones mapping near the telomeres of each chromosome (Knight *et al.*, 2000) and clones containing certain named genes from J. Flint, D. Ledbetter, C. Lese and Vysis, Inc. We included P1 clones used previously on arrays of chromosome 20 (Pinkel *et al.*, 1998; Albertson *et al.*, 2000). We also used additional clones from

the DuPont A library, including RMC01P052 (163D7), RMC01P057 (213C4), RMC07P014 (548H7), RMC07P025 (1128F4), RMC07P028 (252B4) and RMC07P038 (1429E1.c2), as well as RMC17P041 (75D7), RMC17P069 (88H6) and RMCXP002 (124A3) from the DuPont B library. The array provides only minimal coverage of the Y chromosome, since only one clone unique to the Y chromosome is included on the array in addition to the telomeric clones that are shared between the X and Y chromosomes.

### **A.1.3 Isolation of BAC/P1 DNA**

We inoculated 0.1  $\mu$ l of a glycerol stock of BAC or P1 containing bacteria into 5 ml LB media containing chloramphenicol (10  $\mu$ g/ml) or kanamycin (50  $\mu$ g/ml) for BACs or P1s, respectively and incubated the cultures overnight (~16 h) at 37°C with agitation at 225 rpm. We prepared larger cultures for DNA isolation, by inoculating 25 ml LB media, containing the appropriate antibiotic with 200  $\mu$ l of the previously grown overnight culture and incubating in a shaking incubator at 37°C at 225 rpm overnight (~16 h). We monitored bacterial growth by measuring the OD of a 1:10 dilution, which preferably ranged between 0.25 and 0.35 at a wavelength of 600 nm. We used the Qiagen Plasmid Mini kit to isolate DNA, following a modified version of the Qiagen Plasmid Mini Purification protocol. We transferred the 25 ml cultures into 50 ml tubes and centrifuged them for 15 min at 4000 rcf at 4°C to pellet the bacteria. We re-suspended the pellet in Qiagen buffer P1 (1.5 ml), containing RNase A at the manufacturer's recommended concentration and then added buffer P2 (1.5 ml). We mixed the tubes extremely gently by inversion and incubated at room temperature for 5 min. Then, we added buffer P3 (1.5 ml), again mixed very gently by inversion and incubated on ice for 10 min. We inverted each tube once and centrifuged at 4000 rpm at 4°C until the supernatant became clear (45 to 60 min). We filtered the supernatant through a 35-micron nylon mesh prior to loading onto Qiagen-tip 20 columns, which had been equilibrated following the manufacturer's protocol. We followed the manufacturer's protocol for the DNA isolation steps with the exception that the elution buffer QF was heated to 65°C before it was added to the column. After elution, we added 0.56 ml isopropanol to the DNA and incubated overnight at 4°C. We collected the DNA pellet by centrifugation for 45 min at 14000 rpm at 4°C. After aspirating the supernatant, we allowed the pellet to dry in air for at most 30 min before re-suspending the DNA in 50  $\mu$ l of H<sub>2</sub>O. We determined the DNA concentration using a fluorometer. The yield typically ranged between 60-100 ng/ $\mu$ l (i.e. 3-5  $\mu$ g per isolation). To ascertain DNA purity, we digested each BAC (200 ng) with *Hind*III and electrophoresed the digest through a 0.75% agarose gel. We discarded DNA preparations that showed a significant contamination with host bacterial DNA, seen as a background smear of degraded DNA in the gel.

### **A.1.4 Preparation of BAC/P1 DNA representations by ligation-mediated PCR**

We digested the DNA with *Mse*I by incubating overnight at 37°C in a 5  $\mu$ l reaction containing 1.5  $\mu$ l DNA (20 to 600 ng), 0.4 U/ $\mu$ l *Mse*I (New England Biolabs), and 0.4x One-Phor-All-Buffer-Plus (Amersham). For ligation of adapters, we diluted 1  $\mu$ l of each digest to a final concentration of 1 ng/ $\mu$ l with H<sub>2</sub>O and then mixed 1 ng of the digested DNA with 0.5  $\mu$ l One-Phor-All-Buffer-Plus (10x, Amersham), 0.5  $\mu$ l of 100  $\mu$ M Primer 1 (TAACTAGCATGC), 0.5  $\mu$ l of 100  $\mu$ M Primer 2 (5' aminolinker AGTGGGATTCCGCATGCTAGT) and 5.5  $\mu$ l H<sub>2</sub>O. We incubated the mixture at 65°C for 1 min after which we ramped the temperature down



to 15°C with a ramp-speed of 1.3°C per minute. When the temperature reached 15°C, we added 1  $\mu$ l of 10 mM ATP and 1  $\mu$ l T4-DNA ligase (5 U/ $\mu$ l, Gibco BRL) and continued incubation at 15°C overnight. To initiate the first round of PCR amplification, we added a 40  $\mu$ l mixture, consisting of 3  $\mu$ l PCR Buffer 1 (Expand Long Template PCR System, Roche), 2  $\mu$ l of a mixture of each nucleotide (10 mM) and 35  $\mu$ l H<sub>2</sub>O to the ligation reaction. Before we started the PCR program, we incubated the reaction at 68°C for 4 min and then added 1  $\mu$ l DNA polymerase (3.5 U/ $\mu$ l, Expand Long Template PCR System, Roche). We carried out thermal cycling in a Perkin-Elmer Gene Amp PCR System 9700 block as follows: 94°C for 40 s, 57°C for 30 s, 68°C for 1.25 min for 14 cycles, followed by, 94°C for 40 s, 57°C for 30 s, 68°C for 1.75 min for 34 cycles and 94°C for 40 s, 57°C for 30 s and 68°C for 5 min for the final cycle. We electrophoresed 3.5  $\mu$ l of the PCR product through a 1% agarose gel to determine the size range of the amplified DNA, which ideally ranged from 100 to 2000 bp. To make the DNA for spotting on the arrays, we carried out a second round of amplification in a 100  $\mu$ l reaction containing 1  $\mu$ l of the primary PCR product, 4  $\mu$ M Primer 2, TAQ-buffer II (1x; Perkin Elmer), 0.2 mM dNTP mix, 5.5 mM MgCl<sub>2</sub> (Perkin Elmer), 2.5 U Amplitaq Gold (Perkin Elmer) and H<sub>2</sub>O. We carried out an initial incubation at 95°C for 10 min in a MJ Research Peltier Thermal Cycler 225, followed by 95°C for 30 s, 50°C for 30 s and 72°C for 2 min for 45 cycles and finally 7 min at 72°C. This reaction yields  $\sim$ 10  $\mu$ g of DNA, with each fragment containing a 5' amino linker.

### **A.1.5 Preparation of DNA spotting solutions**

We evaporated the amplification reaction (100  $\mu$ l) to a final volume of 50  $\mu$ l by incubation at 45°C in a hybridization oven (Techne, Hybridizer HB-1D) for approximately 75 min and then added 2.5 volumes of ice-cold ethanol and 0.1 volumes of 3 M NaOAc to precipitate the DNA. (The use of ethanol precipitation proved superior to isopropanol precipitation.) We inverted the tubes and chilled them at -20°C for 15 min before collecting the precipitate by centrifugation at 1699 rcf for 90 min. We washed the pellets with 70% ethanol (150  $\mu$ l) and collected them by centrifugation at 1699 rcf for 45 min. We air-dried the pellets for approximately 60 to 90 min and then re-suspended them in 12  $\mu$ l of 20% DMSO in H<sub>2</sub>O ( $\sim$ 0.8  $\mu$ g/ $\mu$ l). We transferred the DNA solutions into 864 well microtitre plates for robotic arraying. Previously, we prepared DNA for spotting in 80% DMSO and 0.3  $\mu$ g/ $\mu$ l nitrocellulose. However, we subsequently evaluated spotting solutions containing various concentrations of dimethyl formamide, formamide or DMSO, with or without nitrocellulose. We determined that nitrocellulose was not required for spotting DNA prepared by PCR using primers with 5' amino linkers and that spotting solutions made with 20% DMSO performed well. We have also generated representations of large insert clones for arraying by using degenerate oligonucleotide primed PCR and by amplification of mixtures of subclones from BACs, but found the ligation-mediated PCR procedure to be superior.

### **A.1.6 Array printing**

We used a custom built printer, employing a 4 x 4 array of quartz capillary tubes spaced on 3 mm centers to print  $\sim$ 70-100  $\mu$ m diameter spots on 130  $\mu$ m centers. We printed each DNA solution in triplicate to create an array of  $\sim$ 7500 elements in a 12 mm square area. We printed the arrays on chromium coated microscope slides (PTI or Nanofilm) for these studies, but also routinely print the arrays on glass slides (Corning GAPS). We allowed the printed

slides to dry overnight, then exposed the slides to UV light (65 mJ) in a UV Stratalinker 2400 (Stratagene). We hybridized to these slides without additional treatment, except for the pre-hybridization slide blocking described below. We found no indication of DNA loss from the spots at any stage of the procedure when we performed hybridization in formamide buffers at 37°C.

### **A.1.7 DNA labeling**

We labeled DNA by nick translation or random priming. For nick translation, we used a 400  $\mu$ l reaction containing 10  $\mu$ g DNA, 50 mM Tris, pH 7.6, 5 mM MgCl<sub>2</sub>, 0.02 mM each dATP, dTTP and dGTP, 0.2 mM Cy3- or Cy5-dCTP (Amersham), 0.2 U DNA Polymerase I (Gibco BRL), and 30  $\mu$ l DNaseI/PolI enzyme mix (Gibco BRL). We incubated the reaction at 15°C for 60 min after which we stopped the reaction by incubation at 70°C for 10 min. We removed unincorporated nucleotides using a Sephadex G-50 spin column. We labeled genomic DNA by random priming in a 100  $\mu$ l reaction containing 0.003 - 0.6  $\mu$ g DNA, 1x random primers solution (BioPrime DNA Labeling System, Gibco BRL), 1 mM Tris, pH 7.6, 0.1 mM EDTA, 0.2 mM each of dATP, dTTP and dGTP, 0.1 mM dCTP, 0.4 mM Cy3 or Cy5-dCTP (Amersham) and 160 U Klenow fragment (BioPrime DNA Labeling System, Gibco BRL). We incubated the DNA with the random primers solution at 100°C for 10 min in a total volume of 84  $\mu$ l, prior to adding the other reagents and then incubated the 100  $\mu$ l reaction overnight at 37°C. We removed unincorporated nucleotides using a Sephadex G-50 column.

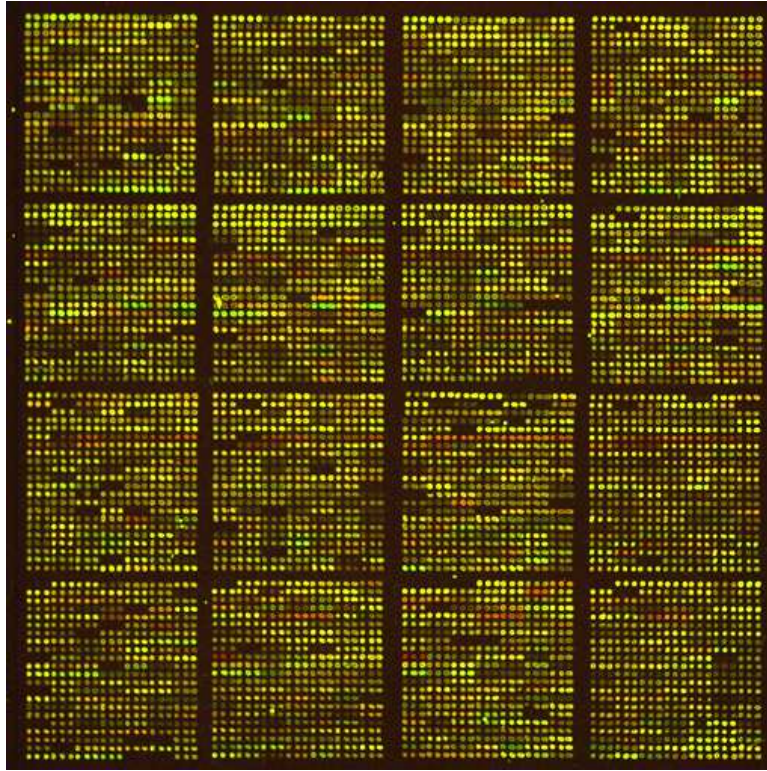
### **A.1.8 Hybridization**

We combined test and reference DNAs ( $\sim$ 2  $\mu$ g of input genomic DNA for nick translation or  $\sim$ 0.6  $\mu$ g of input genomic DNA for random prime labeling) with Cot-1 DNA (80-100  $\mu$ g; Gibco BRL) and precipitated them with ethanol. We collected the precipitate by centrifugation and allowed the precipitate to dry in air for 10 min before re-dissolving it in a 50  $\mu$ l hybridization mixture containing 50% formamide, 2 x SSC, 10% dextran sulfate, 4% SDS and 500  $\mu$ g yeast tRNA, pH 7. We incubated the hybridization mixture at 70°C for 10-15 min to denature the DNA and subsequently continued incubation at 37°C for 60 min to allow blocking of repetitive sequences. We applied a ring of rubber cement closely around the array to form a well, into which we added 50  $\mu$ l of slide blocking solution containing 500  $\mu$ g salmon sperm DNA in 50% formamide, 2 x SSC, 10% dextran sulfate and 4% SDS, pH 7. We created an airtight hybridization chamber by placing a silicone gasket (PGC Scientific) around the array and rubber cement ring, placing a slide on top of it and clamping with binder clips. After a 30 min incubation at room temperature, we opened the chamber, removed approximately three-quarters of the blocking solution, added the denatured and re-annealed hybridization mixture and then re-sealed the chamber. We placed the arrays on a slowly rocking table ( $\sim$ 1 rpm) at 37°C to allow hybridization to occur over 16-72 h. After hybridization, we rinsed off the excess hybridization fluid with PN buffer (PN: 0.1 M sodium phosphate, 0.1% nonidet P40, pH 8), then washed once in 50% formamide, 2 x SSC, pH 7 at 45°C for 15 min, and finally in PN buffer at room temperature for 15 min. After draining excess liquid from the arrays, we mounted them in a solution containing 90% glycerol, 10% PBS and 1  $\mu$ M DAPI, and then sealed them with a cover slip. We drained excess glycerol-DAPI solution onto a paper tissue.

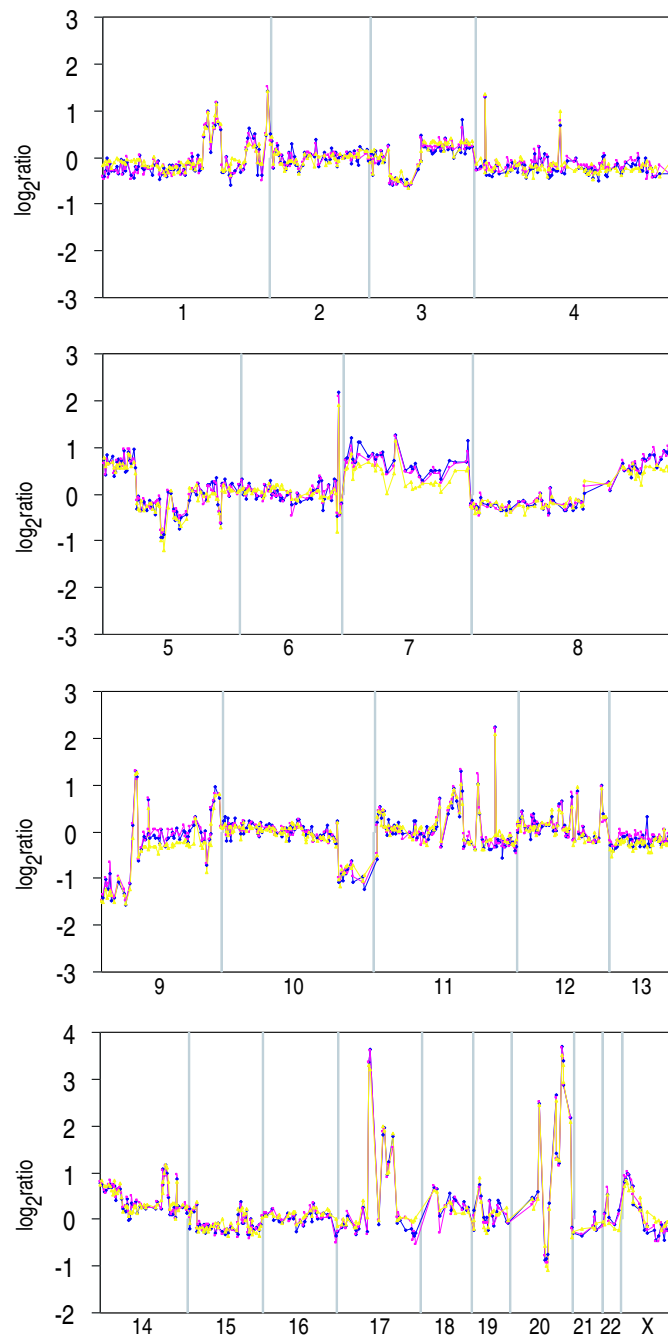
### **A.1.9 Imaging and analysis**

We acquired 16 bit 1024x1024 pixel DAPI, Cy3 and Cy5 images using a custom built CCD camera system (Pinkel *et al.*, 1998), although we have also used commercial laser scanners for this purpose. We used "UCSF SPOT" software (Jain *et al.*, 2001) to automatically segment the spots based on the DAPI images, perform local background correction and to calculate various measurement parameters, including  $\log_2$  ratios of the total integrated Cy3 and Cy5 intensities for each spot. We used a second custom program SPROC to associate clone identities and a mapping information file with each spot so that the data could be plotted relative to the position of the BACs on the September, 2000 freeze of the draft human genome sequence (<http://genome.ucsc.edu>). SPROC also implements a filtering procedure to reject data based on a number of criteria, including low reference/DAPI signal intensity and low correlation of the Cy3 and Cy5 intensities with a spot. The SPROC output consists of averaged ratios of the triplicate spots for each clone, standard deviations of the triplicates and plotting position for each clone on the array, as well as other clone information stored in the database, such as STS content (Web Table A). We edited the data files to remove ratios on clones for which only one of the triplicates remained after SPROC analysis and/or the standard deviation of the  $\log_2$  ratios of the triplicates was  $> 0.2$ .

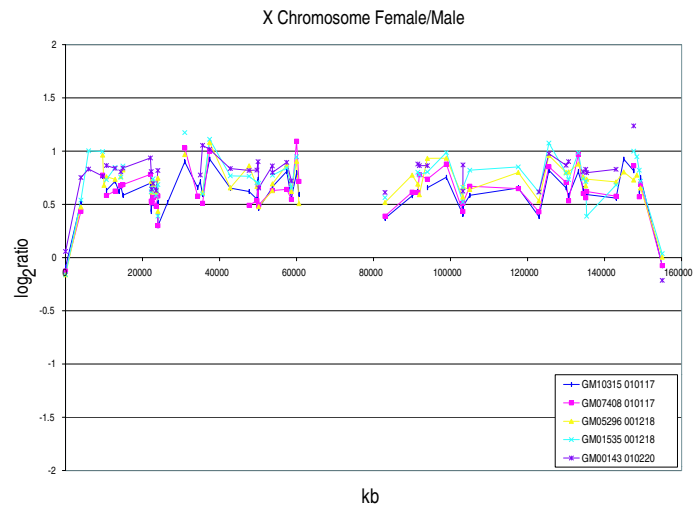
## A.2 Supplementary Figures and Table



**Figure A.1:** We hybridized Cy3 labeled HT29 cell line DNA (green) and Cy5 labeled normal male reference DNA (red) to an array and imaged it using the CCD camera system. We produced the two color image by overlaying the individual Cy3 and Cy5 images using Adobe Photoshop.



**Figure A.2:** Normalized copy number ratios for three comparisons of BT474 DNA and normal male reference DNA on an array of 1777 clones (HumArray 1.11). In two hybridizations, we labeled the genomic DNAs by random priming (pink and blue) and in the third by nick translation (yellow). Data are plotted as the mean log<sub>2</sub>ratio of the triplicate spots for each clone normalized to the genome median log<sub>2</sub>ratio. The BACs are ordered by position in the genome beginning at 1p and ending at Xq. The borders between chromosomes are indicated with vertical bars.



**Figure A.3:** Normalized copy number ratios for X chromosome clones for 5 female/male comparisons. We found consistent ratio differences among the clones. Data are plotted as the mean  $\log_2$  ratio of the triplicate spots for each clone normalized to the genome median  $\log_2$  ratio. The BACs are ordered by position from Xp<sub>ter</sub> to Xq<sub>ter</sub>. Note that ratios on the Xp and Xq telomere clones are at the genome average, since these clones contain sequence shared with the Y chromosome.

**Table A.1:** Summary of genome-wide copy number variation in aneuploid cell strains.

Cell line	Karyotype <sup>1</sup>	Proximal flanking clone	Proximal flanking STS <sup>2</sup>	Proximal included clone	Proximal included STS	Distal included clone	Distal included STS	Distal flanking clone	Distal flanking STS
GM03563	46,XY,-9,+der(9)t(3;9)(qter>9p24::5q12>3qter)mat Trisomy 3q12-3qter Monosomy 9pter-9p24	RP11-146e16 None	AFMA059XB5	CTD-2014B13 CTB-41L13	stSG9229 9 p tel	GS1-56H22 RP11-28N06	3 q tel SHGC-1302	None RP11-122A20	SHGC-12535
GM00143	47,XX,+18 Trisomy 18	None		GS1-74G18	18 p tel	RP11-507p3	18 q tel	None	
GM06296	46,XX,-11,+der(11)inv ins(11;10)(11pter>11p13::10q21>10q24::11p13>11qter)mat Trisomy 10q21-10q24 Monosomy 11p12-11p13	RP11-237j07 CTD-2208j5	SHGC-17478 X51630	RP11-33a02 RP11-127K23	SHGC-10261 SHGC-3931	RP11-128h11 RP11-72A10	SHGC-2001 SHGC-20671	RP11-46g13 RP11-18B09	SHGC-20638 AFM270VB1
GM07408	47,XX,+20 Trisomy 20	None		RP1-82O2	20 p tel	RP1-81F12	20 q tel	None	
GM01750	47,XY,+der(14)t(9;14)(14pter>14q21::9p24>9pter)mat Trisomy 9pter-9p24 Trisomy 14pter-14q21	None None		CTB-41L13 RP11-14a20	9 p tel SHGC-11728	RP11-33O15 RP11-125A05	SHGC-8727 SHGC-1403	RP11-85J05 RP11-48L01	SHGC-844 SHGC-6471
GM03134	46,XY,del(8)(pter>q13;q22>qter) Monosomy 8q13-8q22	RP11-117N14	SHGC-586	RP11-33D07	SHGC-13825	RP11-271f5	SHGC-5171	RP11-102K07	SHGC-12502
GM13330	46,XY,-4,+der(4)t(1;4)(4pter>4q35::1q25>1qter)mat Trisomy 1q25-1qter Monosomy 4q35-4qter	RP11-234m03 RP11-272C03	AFMA129ZF5 SHGC-4789	RP11-28d10 RP11-244K02	AFM063XG9 SHGC-4-548	GS1-167K11 GS1-31J3	1 q tel 4 q tel	None None	
GM03576	48,XY,+2,+21 Trisomy 2 Trisomy 21	None None		GS1-8L3 RP11-31b06	2 p tel AFM289YH9	CTB-172113 RP11-135b17	2 q tel 21 q tel	None None	
GM01535	46,XX,-12,+der(12)t(5;12)(12pter>12q24::5q33>5qter)mat Trisomy 5q33-5qter Monosomy 12q24-12qter	RP11-210k16 RP11-81g12	AFMA102ZG9 SHGC-372	RP11-88j19 RP1-221K18	AFM211YC7 12 q tel	RP1-240g13 None	5 q tel None	None None	
GM07061	46,XY,-15,+der(7)t(7;15)(7pter>7q11.2::15q11.2>15qter)mat Trisomy 7pter-7q11.2 Monosomy 15pter-15q11.2	None not detected		RP1-164D18	7 p tel	RP11-251115	AFM191XH6	RP11-90O18	SHGC-33722
GM02948	47,XY,+13 Trisomy 13	None		RP11-8c15	SHGC-37580	RP1-01L16	13 q tel	None	
GM04435	48,XY,+16,+21 Trisomy 16 Trisomy 21	None None		CTD-2371a5 RP11-31b06	16 p tel AFM289YH9	PAC 191p24 RP11-135b17	16 q tel 21 q tel	None None	
GM10315	47,XX,+22 Trisomy 22	None		RP1-1516	D22S543	CTA-799f10	22 q tel	None	
GM13031	46,XY,del(17)(pter>q21.3::q23>qter) Monosomy 17q21.3-17q23	RPS-1071f14	D17S797	RP11-481c4	D17S1989	RP11-67O13	WI-9306	RP11-50F16	SHGC-34099
GM01524	46,XY,-5,+der(5)ins(5;6)(5pter>5q33::6q15-6q25::5q33>5qter)mat Trisomy 6q15-6q25	CTD-2009c06	stSG6355	RP11-2m09	AFM295TB3	RP11-107m03	AFM198WG11	RP11-139e22	AFM059YD6

<sup>1</sup>Karyotype as proposed by Coriell

<sup>2</sup>STS = sequence tagged site

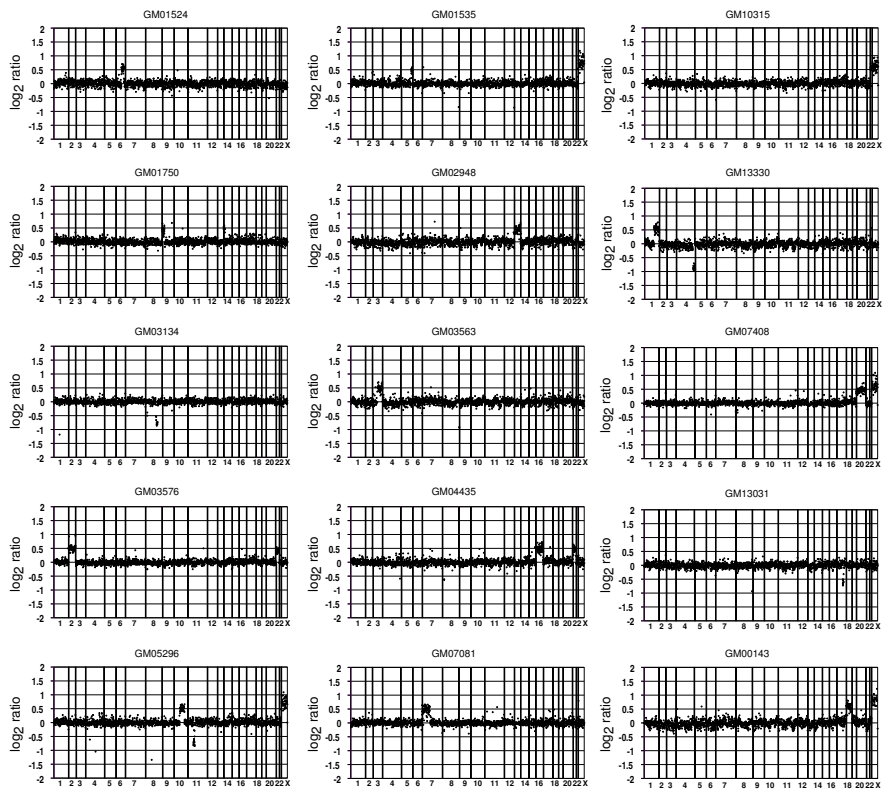


Figure A.4: Normalized copy number ratios for Coriell cell strains.



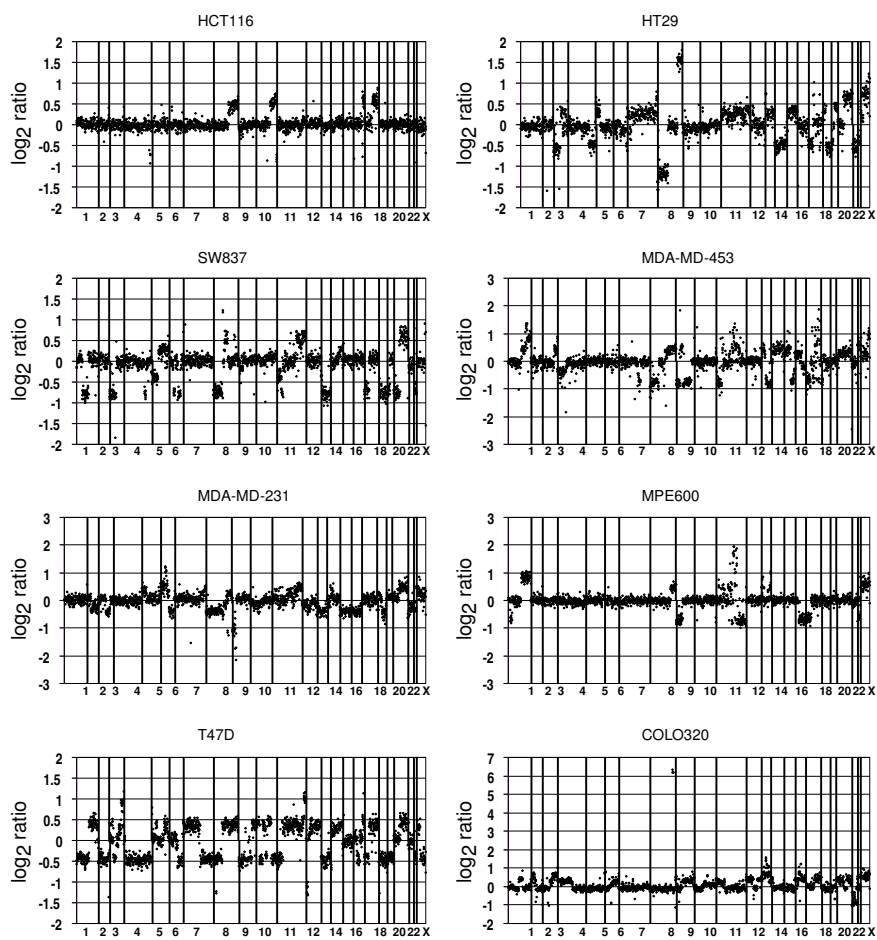
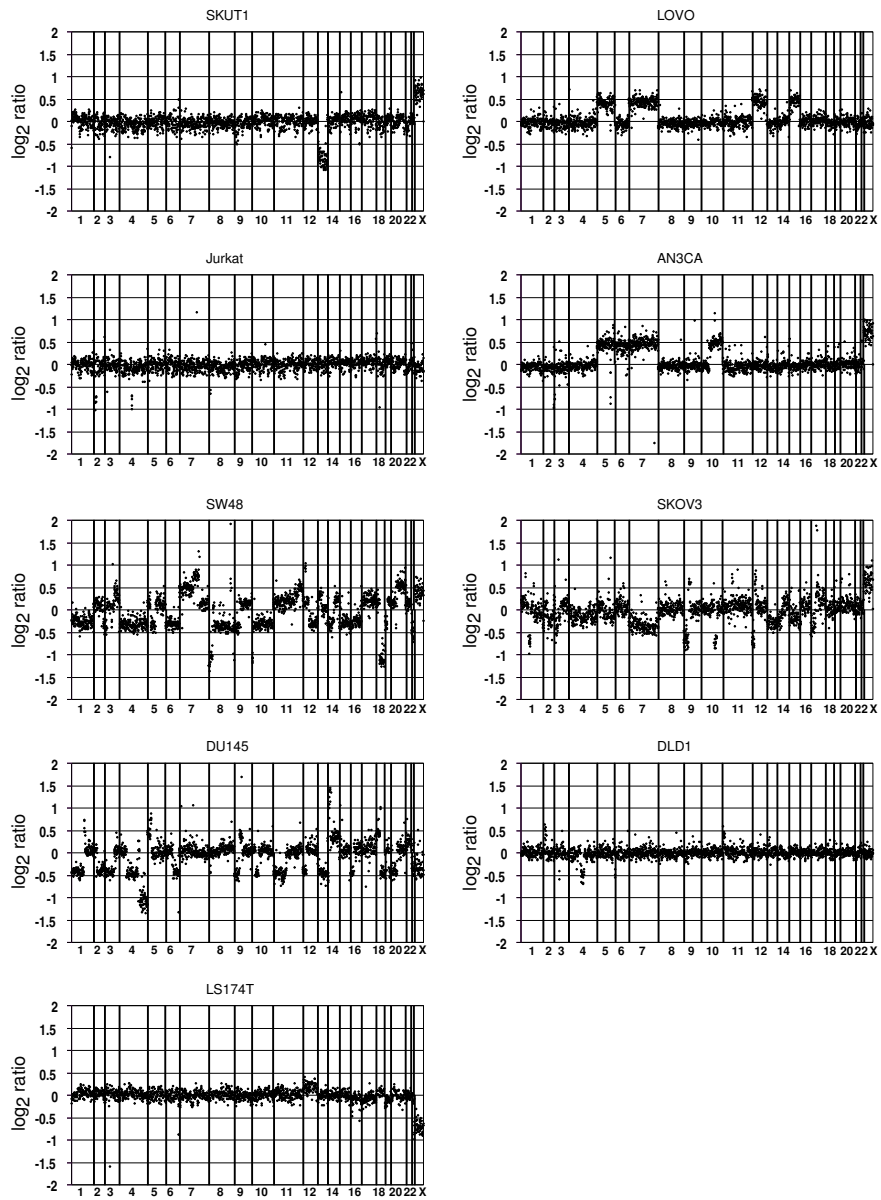


Figure A.5: Normalized copy number ratios for cancer cell lines.

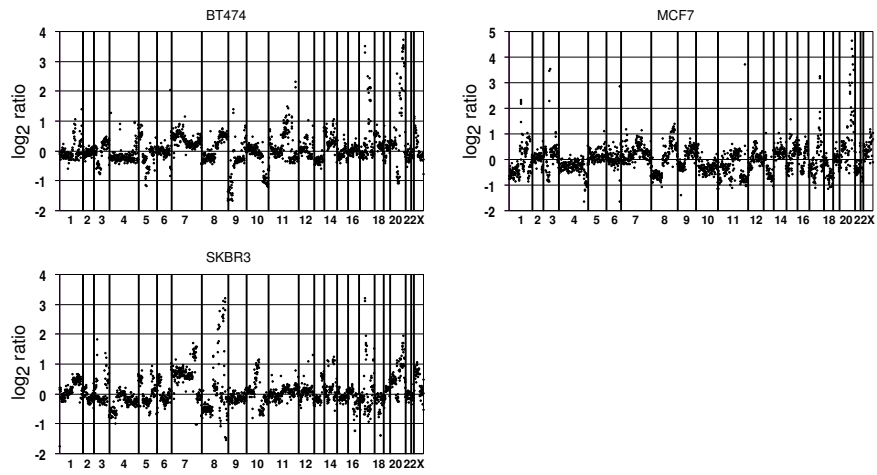


## **Appendix B**

# **Supplementary information: Shaping of Tumor and Drug Resistant Genomes by Instability and Selection**



**Figure B.1:** Normalized copy number ratios for mismatch repair deficient cancer cell lines.

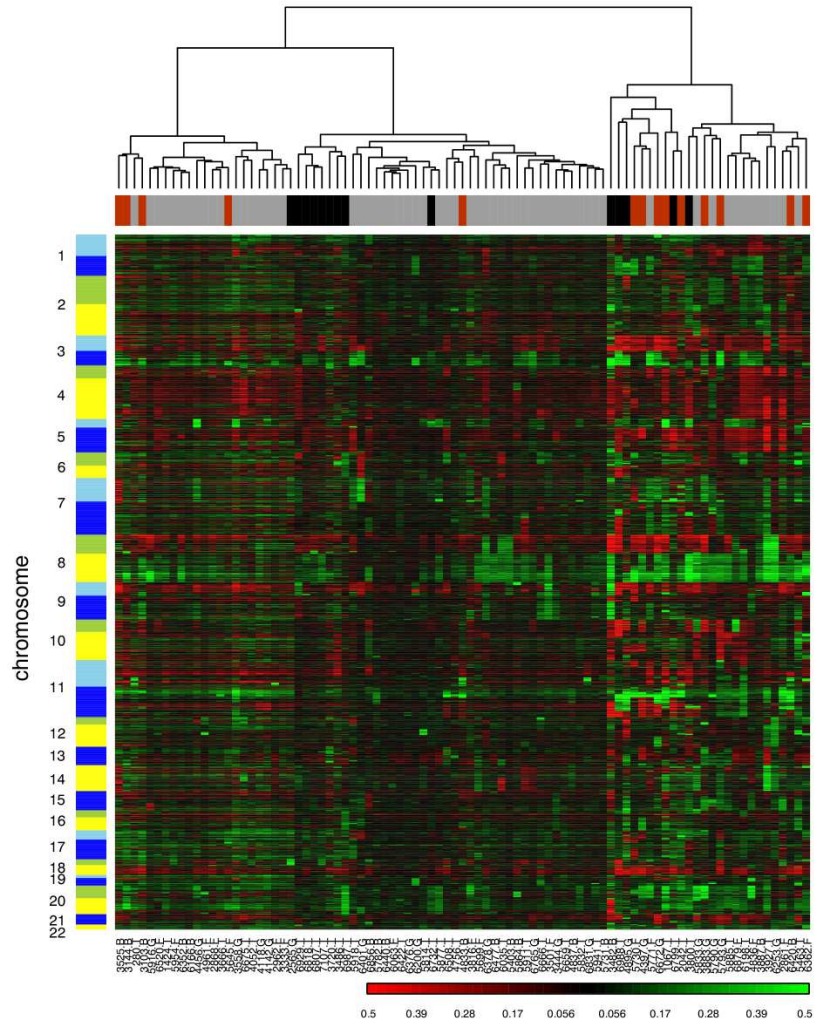


**Figure B.2:** Normalized copy number ratios for mismatch repair proficient cancer cell lines.



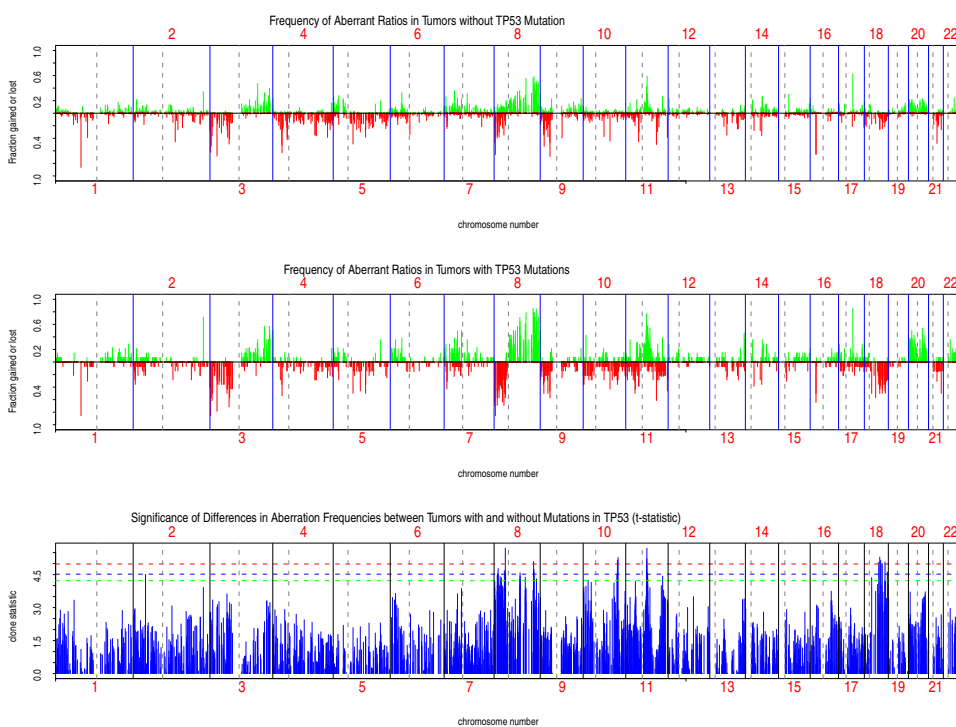
## **Appendix C**

# **Supplementary information: Rare Amplicons Implicate Frequent Misspecification of Cell Fate in Oral Squamous Cell Carcinoma**



**Figure C.1:** Hierarchical clustering of 89 oral SCC based on their genome-wide DNA copy number profile. We represented individual clones as rows and ordered them by chromosome and genome position according to the July 2003 freeze of the human genome. We indicated clones on the p-arm either in light blue or yellow, and clones on the q-arm in dark blue or green. We show acrocentric chromosomes in yellow or dark blue. Columns represent individual tumor samples. We indicate the *TP53* mutation status of tumors with a red box for *TP53* mutant tumors, a gray box for tumors with no detected mutation and a black box if the *TP53* status is unknown. The cluster on the far right has a significantly higher proportion of *TP53* mutant tumors, suggesting an association of *TP53* mutational status with particular DNA copy number aberrations in oral SCC. We found that 18% of oral SCCs harbored a mutation in exons 5-8 of *TP53* (5 buccal mucosa, 3 floor of mouth, 4 gingiva and 2 tongue).





**Figure C.2:** Frequency plots of oral SCC with and without mutation in *TP53* exons 5-8. The top panels show the frequency with which we found aberrant ratios on individual clones in at least 10% of samples. For each clone, we show gains (indicated in green, ranging from 0 to 1) and losses (indicated in red, ranging from 0 to -1) using tumor specific thresholds for 89 SCCs sorted according to *TP53* mutation status. The bottom panel shows the level of significance of the difference (t-statistic) between the two sets of tumors at each clone. The dashed horizontal lines indicate  $p=0.01$  (green),  $p=0.05$  (blue) and  $p=0.001$  (red).



# Summary and discussion

Changes in DNA copy number are frequently observed in many types of cancer, developmental abnormalities and when comparing different populations, strains or species. Detecting these DNA copy number variations is useful in understanding complex diseases such as cancer. This knowledge can also be used to identify potential targets for the development of therapeutics; it can be used as a diagnostic tool or help us understand the evolutionary history of different species. This thesis describes both the development and applications of genome-wide array comparative genomic hybridization to detect DNA copy number variations in both human and mouse genomes.

Large clones, such as bacterial artificial chromosomes (BACs), containing  $\sim 150$  kb of genomic DNA insert provide sufficient signal intensity to quantitatively detect single copy number changes as well as homozygous deletions and high-level amplifications. However, BACs are single copy vectors and the yield of standard BAC preparations is relatively low. Moreover, accurately depositing whole BAC DNA, dissolved in a small volume, on a glass substrate is problematic due to the viscosity of the solution. Therefore, we explored different means to amplify and reduce the size of BAC DNA, while at the same time maximizing the representation of each BAC in the spotting solution. First we tried to reduce the physical size of BAC DNA by sonication. This randomly shears DNA, rendering it less viscous after preparation of the DNA spotting solution. Although the measured ratios after hybridization were comparable to ratios obtained from arrays containing whole BAC DNA, and the viscosity issue of whole BAC DNA was solved, problems remained regarding the production of sufficient BAC DNA for the production of genome-wide human and mouse arrays. In a second approach, we sonicated BAC DNA and treated the DNA with either an exonuclease or a polymerase rendering the ends of the DNA blunt. The blunt-ended BAC DNA fragments were subsequently cloned into a common vector. Bacteria were transformed using this pool of cloned BAC DNA fragments. A single polymerase chain reaction (PCR) using a standard primer pair, which is present in the cloning vector would in theory yield sufficient BAC DNA for arraying while at the same time maintain the representation of the original BAC DNA. Although, sufficient BAC DNA was obtained after PCR, the ratios after hybridization were inconsistent and differed from ratios obtained with arrays containing whole BAC DNA. In Chapter 3 we describe our third and most successful approach. Here we adapted a ligation-mediated PCR method, originally developed for the representational amplification of whole genomic DNA. BAC DNA is first digested with a frequently cutting restriction enzyme, leaving 2 basepair overhangs on either side of the fragments. A partially overlapping oligonucleotide pair is ligated to 1 ng of the digested DNA. A primary PCR is carried out using one of the two oligonucleotides as a primer. Using the same primers a secondary PCR

is carried out using only 1  $\mu$ l of the primary PCR in a larger volume. This PCR is processed into spotting solution. Since only 1  $\mu$ l of the primary PCR is used for each secondary PCR and since each secondary PCR provides enough spotting solution for thousands of arrays, this method basically provides an unlimited supply of BAC DNA spotting solution. Also, the ratios obtained after hybridization from arrays containing spotting solutions prepared by ligation-mediated PCR are almost identical to ratios obtained from arrays containing whole BAC DNA. It should be stressed that using these arrays we can quantitatively detect low-level (i.e. single copy number aberrations) as well as high-level copy number aberrations on a single arrayed element (i.e. a single BAC). We could not only differentiate different types of aberrations but also quantify the number of different types of aberrations seen in tumors and different mouse strains. Using the ligation-mediated PCR method we generated a human array containing 2464 BACs and a mouse array containing  $\sim$ 2500 BACs. On the human array, each BAC maps to a single site by FISH and contains at least one STS, which links each BAC to the sequence of the human genome. Each BAC on the mouse array also contains at least one STS. Although multi-site BACs on these arrays are flagged, they may still be of interest when investigating DNA copy number polymorphisms between different strains or species. In chapter 4 we show that different mouse strains are indeed characterized by the presence of multiple DNA copy number polymorphisms. We identified 75 polymorphic regions (80 polymorphic BAC clones) in seven inbred *Mus musculus* strains as well as one inbred and one outbred *Mus spretus* strain. The majority of these variant loci are recurrent among different individuals from one strain and all analyzed strains can be distinguished based on these loci. Using fluorescent *in situ* hybridization (FISH) we determined that out of 80 polymorphic BAC clones 63 hybridized to a single site while 17 hybridized to multiple places in the genome. We observed that a number of these multi-site BACs hybridized to similar chromosomal locations. Specifically, loci on chromosomes 1 (distal), 5 (middle), 7 (proximal), 9 (proximal and middle), 12 (proximal and middle), 14 (proximal) and 17 (proximal/middle) more frequently hybridized with these BACs. Also several multi-site BACs hybridized to more than one of these loci. Most of the multi-site BACs hybridized more frequently to the proximal 1/3 of the chromosomes, near the pericentromeric regions of the chromosomes in addition to a site more distal on the same or other chromosome(s). This is consistent with the previously reported distribution of segmental duplications throughout the human and mouse genomes. No such bias was observed between the BAC clones that hybridized to a single site. It is likely that these regions harbor gene families or low copy number repeats thereby increasing the possibility of unequal crossing-over resulting in inter-strain sequence differences. However, more research is needed to confirm this, to identify the sequence content of these regions and to study the functional consequences of these copy number variations.

We also introduced a new application for array based CGH by showing that we could differentiate regions of heterozygosity in different backcross NIHxSPRET/Glasgow mice. Regions homozygous for NIH showed a subtle increase in  $\log_2$ ratio while regions heterozygous for NIH showed a subtle decrease in  $\log_2$ ratio after hybridization using NIH as a reference. We propose that the observed increase in ratio in regions homozygous for NIH is due to a more perfect match between NIH genomic DNA and the C57Bl/6 BAC DNA in the array spots. Similarly, the observed decrease in ratio in regions heterozygous for NIH is due to a relatively less efficient hybridization of SPRET/Glasgow DNA to the C57Bl/6 BAC DNA compared to NIH. Since the here described array CGH technology is optimized for the detection of DNA copy number aberrations and not sequence divergences between related strains and species, it is foreseeable that a more pronounced ratio difference can be obtained by adjusting the nec-

essary parameters. The two main advantages of using an array CGH based approach for the detection of regions of heterozygosity are the relatively large number of markers across the genome which are genotyped in a single experiment and the speed at which many different mice can be processed at one time compared to conventional methods. Whether this platform is sensitive enough to detect regions of heterozygosity between backcrosses produced by crosses of closely related inbred strains remains to be seen.

The relationship between DNA copy number aberrations and underlying genetic defect is poorly understood for the majority of solid tumors. Mismatch repair (MMR) deficient tumors are probably the most well characterized example. These tumors show a high level of microsatellite instability due to defects in mismatch repair genes (predominantly *MSH2*, *MSH6* and *MLH1*). Mismatch repair deficient cells accumulate small nucleotide changes in regions that are more prone to errors resulting from the absence of this type of DNA repair and those that promote cell growth and/or increase survival will be the ones observed in the tumor. The spectrum of aberrations in MMR deficient tumor genomes will then be a reflection of selection for alterations that promote cell growth and/or increase survival as well as the underlying genetic defect that promotes certain types of changes in the genome. In line with this are observations that MMR deficient tumors are characterized by the presence of few chromosome level aberrations. In chapter 5 we confirm the presence of few chromosome level aberrations in MMR deficient cell lines when compared to MMR proficient cell lines. Interestingly, *MSH2/6* deficient cells harbor fewer copy number transitions and focal aberrations when compared to *MLH1* deficient cells. To extend this study of determining the role of the genetic background on the types and numbers of DNA copy number aberrations we used a model system where mismatch repair (MMR) deficient and proficient cells were selected for methotrexate (MTX) resistance. After selection, the types and numbers of copy number aberrations were analyzed by array CGH. This study showed that different aberrations were associated with MMR status, specifically, the MMR proficient cell line frequently showed amplification while the MMR deficient cell line frequently showed no copy number aberrations after selection for MTX resistance.

Fallopian tube carcinoma (FTC) is a rare and aggressive cancer. We analyzed copy number aberrations in 14 FTC using genome-wide array CGH. The results are discussed in chapter 6. All tumors showed a high level of DNA copy number aberrations. Several known oncogenes map to recurrent regions of amplifications, including *CCNE1*, *MYC* and *AKT2*. We also observed that these tumors are very homogeneous with some aberrations occurring in more than 70% of samples.

In chapter 7 we provide evidence that second primary tumors that arise in the oral cavity are clonally related to the index tumor, which sometimes occurred at an anatomically distinct site and/or more than a decade prior to the presentation of the second tumor in some cases. Using array CGH we compared multiple tumors from each patient in a total of three patients. Although we observed tumor specific aberrations in almost all tumors, it was clear that these tumors have a common origin based on the presence of so called "signature" copy number aberrations. This work provides evidence for the possibility that tumor cells from the index tumor sloughed off and migrated through the oral cavity and seeded elsewhere or that both tumors arose from a field of clonally related cells containing specific aberrations.

Oral squamous cell carcinoma (oral SCC) has a 5-year survival of only 40%, and has not improved over the past 40 years. To identify pathways involved in the development or maintenance of oral SCC we performed genome-wide array CGH on 89 SCC that arose in different sites within the oral cavity. The results are discussed in chapter 8. We identified nine recurrent regions of amplification smaller than 3 Mb in size. Two are frequently amplified in oral SCC and are located on 7p11.2 encompassing *EGFR* and on 11q13 encompassing *CCND1*, *FGFR3*, *FGFR4*, *EMSI* and, perhaps on a separate amplicon, *PAK1*. Both amplicons are positively correlated with a mutation in exons 5-8 of *TP53*. Seven amplicons, however, occurred less frequently and have not been associated with oral SCC before. To identify driver genes for these amplicons we performed expression analysis on candidate driver genes within these amplicons. We identified genes involved in integrin signaling, survival, adhesion and migration, and, interestingly, Hedgehog and Notch pathway genes. Both Hedgehog and Notch pathways are involved in cell fate decision and differentiation. We performed expression analysis on additional members of both the Hedgehog and Notch pathways, showing deregulation of these pathways, implicating cell fate misspecification in oral SCC development. Given the poor 5-year survival rate and the lack of improvement over the past 40 years, it seems reasonable to explore new therapeutic approaches by targeting these pathways in oral SCC.

In the future, for array CGH to be clinically useful one needs a robust platform that can quantitatively detect a wide dynamic range of copy number variations. Array production needs to be closely monitored and verified to ensure consistent quality across many arrays. Moreover, hybridization and analysis procedures need to be streamlined in such a way that the ultimate diagnosis or prognosis can be unambiguously determined. In this thesis we have introduced an array CGH platform that is capable of detecting a wide range in copy number variations, ranging from homozygous deletions, low-level aberrations and high-level amplifications on a single arrayed element and therefore holds great potential to be introduced in the clinic. One can envision the use of array CGH based assays in many different clinical settings in both medical and tumor genetics. Currently, fluorescent *in situ* hybridization (FISH) and karyotyping are the gold standard for the detection of DNA copy number aberrations in clinical samples. For example, FISH is used to determine DNA copy number levels of *ERBB2*, which is frequently amplified in breast cancer and has been suggested to play a role in breast cancer progression. Amplification of *ERBB2* is also an indicator of poor prognosis. Standard karyotyping and FISH are used to diagnose different developmental abnormalities which are characterized by duplication, deletions, translocations or inversions. Although the use of FISH can improve the resolution by which many of these aberrations can be detected compared to karyotyping, one typically only probes a single locus, which means that prior knowledge about the possible presence of a certain well defined aberration, such as amplification of *ERBB2*, is a prerequisite. Also sample preparation and data analysis are laborious when using karyotyping or FISH. By using genomic arrays one can potentially detect DNA copy number aberrations with increased resolution, map these aberrations directly onto the sequence of the genome and interrogate many different loci in a large number of samples. Tumor genomes are very complex often containing numerous aberrations, which not always result in a change in DNA copy number. It is therefore difficult to design an array CGH assay, which will provide clinically useful data (i.e. containing diagnostic and/or prognostic value) in the majority of tumors. On the other hand, many developmental abnormalities are characterized by known duplications or deletions. It is therefore likely that the first clini-

cal applications of array CGH will screen for the presence of DNA copy number aberrations, which may lead to various dysmorphologies and/or mental retardation, in pre- and postnatally obtained samples. A potential pitfall which needs to be addressed is the possibility that the arrayed elements are polymorphic in the population. In this scenario, depending on which two samples are hybridized one can observe a change in DNA copy number, which is not necessarily the cause of an observed phenotypic appearance.

In chapter 7 we showed that different tumors, separated in time and space, within the oral cavity are clonally related. It is very well possible that these tumors arose from a clonally derived field of aberrant cells. Using array CGH in combination with FISH on biopsies or brushes one should be able to identify the extent of a possible field of aberrant cells which would allow clinicians to more accurately tailor treatment to individual patients.

Although tumor specific array CGH assays for clinical use may be difficult to envision in the near future, the use of array CGH for tumor profiling will still provide interesting data. One can narrow down recurrent regions of DNA copy number variation across many tumors, thereby identifying markers that may ultimately find its way into the clinic. In addition, as shown in chapter 5, array CGH provides an opportunity to study the role of specific gene defects on the types and numbers of copy number aberrations present in the tumor. Different genetic backgrounds may lead to different spectra of genomic aberrations. Ultimately, we may not be able to associate a specific gene defect solely with the shape of a genomic DNA copy number profile, but we might be able to predict defects in certain pathways or mechanisms. Also, clustering tumors based on their copy number profiles might classify tumors which then could be correlated with clinical data even though we might not yet fully understand the underlying defects that gave rise to these profiles.

In conclusion, the field of array CGH has made rapid progress over the last five years. We have introduced a human and a mouse platform for array CGH with which we can accurately determine a wide dynamic range of copy number aberrations. We also presented novel applications for this technology, and we will soon see the translation of array CGH from a research setting towards a clinical setting.





# Samenvatting

Chromosomale afwijkingen worden vaak gedetecteerd in verschillende vormen van kanker en in mensen met ontwikkelingsproblemen. Wanneer verschillende populaties of populaties onderling met elkaar worden vergeleken worden vaak chromosomale verschillen waargenomen. Het detecteren van deze chromosomale afwijkingen of verschillen kan dus inzicht geven in complexe ziektes en in de evolutionaire geschiedenis van verschillende populaties. Deze kennis kan ook gebruikt worden in de ontwikkeling van medicatie en voor diagnostische doeleinden. In dit proefschrift worden zowel de ontwikkeling als toepassingen beschreven van 'genome-wide array comparative genomic hybridization' (array CGH) voor het detecteren van chromosomale afwijkingen en verschillen in het genoom van mens en muis.

Met de techniek 'array CGH' kunnen chromosomale veranderingen in aantallen kopieën van een bepaalde sequentie in het genoom gedetecteerd worden ('gains' en 'losses'). Meestal wordt een test genoom (bijvoorbeeld genomisch DNA van een tumor) en een referentie genoom (meestal een normaal genoom) gelabeld met twee verschillende fluorochromen, samengevoegd en gehybridiseerd op een array van genomische klonen. Na de hybridisatie is de ratio van de intensiteit tussen test- en referentiesample een maat voor het relatieve verschil in aantallen kopieën tussen deze twee genomen. Met behulp van een specialistische array-printer kunnen duizenden klonen dicht naast elkaar geprint worden. Bijvoorbeeld, een array met 3000 klonen die verspreid liggen over het genoom leidt tot een resolutie van ongeveer 1 megabase (Mb). Klonen die een groot segment (~150 kb) genomisch DNA bevatten, zoals bijvoorbeeld BAC (bacterial artificial chromosome) klonen, geven na hybridisatie een voldoende hoge intensiteit, wat van groot belang is voor het detecteren van homozygote en hemizygote deleties en duplicaties. Echter, de opbrengst van een BAC kweek cultuur is relatief laag. Verder is het moeilijk om BAC DNA, dat opgelost is in een klein volume, betrouwbaar en nauwkeurig op een glazen ondergrond te printen. Dit is voornamelijk te wijten aan de viscositeit van de oplossing. Daarom hebben wij verschillende mogelijkheden bestudeerd waarbij we BAC DNA amplificeren, de grootte reduceren en tegelijkertijd de algehele BAC representatie maximaliseren.

In hoofdstuk 3 beschrijven we onze meest succesvolle benadering: ligatie gemedieerde 'polymerase chain reaction' (PCR). Iedere BAC wordt eerst met een frequent knippend restrictie enzym geknipt waarna aan de 'sticky ends' van slechts 1 nanogram van dit geknipt BAC DNA een oligonucleotide paar wordt geligeerd. Voor iedere BAC worden deze DNA fragmenten vervolgens geamplificeerd door middel van PCR waarbij één van de twee oligonucleotiden als primer dient. Een tweede PCR reactie met dezelfde primers en met als template slechts 1  $\mu$ l van de eerste PCR wordt gebruikt voor het maken van de print oplossing. Aangezien

slechts 1  $\mu$ l van de eerste PCR gebruikt wordt voor iedere tweede PCR en aangezien van iedere tweede PCR genoeg printoplossing gemaakt kan worden voor duizenden arrays kan met deze methode een praktisch ongelimiteerde hoeveelheid printoplossing gemaakt worden. Belangrijk hierbij is dat de ratio's na hybridisatie vergelijkbaar waren met ratio's verkregen van arrays die complete BAC klonen bevatten. Met behulp van deze ligatie-gemedieerde PCR techniek hebben we een humane array van 2464 BAC klonen en een muizen array van ongeveer 2500 BAC klonen geproduceerd. Met deze arrays kan dus voor elke geprinte kloon het relatieve verschil in kopieën kwantitatief gemeten worden. Dit verschil in aantallen kopieën kan uiteenlopen van homozygote en hemizygoten deleties tot duplicaties en amplificaties. Iedere BAC kloon heeft tenminste één specifieke marker ('sequence tagged site') waardoor deze op de sequentie van het humane of muizen genoom geplaatst kan worden.

In hoofdstuk 4 bestuderen we de aanwezigheid van duplicatie of deletie polymorfismen in verschillende muizenstammen met array CGH. In zeven ingeteelde *Mus musculus* stammen, één ingeteelde *Mus spretus* en één niet-ingeteelde *Mus spretus* stam hebben we 80 polymorfe BAC klonen ontdekt. Een aantal van deze polymorfe BACs zijn gedeeltelijk identiek waardoor het totale aantal polymorfe loci uitkomt op 75. We laten zien dat verschillende muizenstammen geclusterd kunnen worden gebaseerd op deze polymorfismen. Met een fluorescentie *in situ* hybridisatie (FISH) experiment hebben we vastgesteld dat 17 van deze 80 BACs naar meerdere loci in het genoom hybridiseren ('multi-site'), terwijl 63 BACs naar één locus hybridiseren ('single site'). Een aantal van deze 'multi-site' BACs hybridiseert naar eenzelfde locus op hetzelfde chromosoom. Zo hybridiseren deze BACs frequenter naar chromosoom 1 (distaal), 5 (middel), 7 (proximaal), 9 (proximaal en middel), 12 (proximaal en middel), 14 (proximaal) en 17 (proximaal en middel). Een aantal 'multi-site' BACs hybridiseren tegelijk naar een aantal van deze loci. Ook zagen we een tendens waarbij de 'multi-site' BACs frequenter naar het meest proximale gedeelte van de chromosomen hybridiseerden. Deze observaties zijn consistent met de distributie van 'segmental duplications' in het genoom van mens en muis. De 'single-site' BACs laten geen specifiek hybridisatie patroon zien. Het is aannemelijk dat deze loci genenfamilies of 'low-level repeats' bevatten waardoor een verhoogde kans op ongelijke 'crossing-over' bestaat. Ook beschrijven we in dit hoofdstuk een nieuwe toepassing van array CGH waarbij homozygote loci kunnen worden onderscheiden van heterozygote loci in verschillende teruggekruiste (NIH X SPRET/Glasgow)F1 X NIH muizen. Voor deze hybridisaties is NIH DNA gebruikt als referentie genoom. Homozygote loci voor NIH in de teruggekruiste muizen laten een subtiele verhoging zien in  $\log_2$ ratio terwijl heterozygote loci een subtiele verlaging in  $\log_2$ ratio laten zien. Dit verschil in  $\log_2$ ratio tussen homozygote en heterozygote loci is waarschijnlijk te wijten aan een betere 'match' tussen NIH en het C57Bl/6 BAC DNA in vergelijking met de 'match' tussen SPRET/Glasgow en het C57Bl/6 BAC DNA. Momenteel worden homozygote en heterozygote loci in teruggekruiste muizen in kaart gebracht door stam specifieke markers met behulp van PCR te amplificeren. Meestal worden een paar honderd markers gebruikt in deze studies. De twee voornaamste voordelen van het gebruik van array CGH voor dit doel is de mogelijkheid om ~3000 markers (of 30.000 met een genoom dekkende array) in één experiment te analyseren en de hoge snelheid waarmee tegelijkertijd verschillende muizen gekarakteriseerd kunnen worden.

De relatie tussen verschillende chromosomale afwijkingen en de onderliggende genetische achtergrond is niet goed bekend voor de meeste solide tumoren. Echter, 'mismatch repair'

(MMR) gemuteerde tumoren zijn waarschijnlijk het best gekarakteriseerde voorbeeld. Deze tumoren hebben één of meerdere mutaties in MMR genen wat een verhoogde microsatelliet instabiliteit tot gevolg heeft. Dit leidt ertoe dat in deze tumoren kleine nucleotide veranderingen (mutaties) accumuleren in loci die gevoelig zijn voor microsatelliet instabiliteit. Mutaties die leiden tot verhoogde proliferatie of overleving worden door positieve selectie teruggevonden in de tumor. De afwijkingen die aanwezig zijn in MMR gemuteerde tumoren zijn dus niet alleen een weergave van de positieve selectie voor afwijkingen die leiden tot verhoogde proliferatie of overleving, maar ook een weergave van de onderliggende genetische achtergrond die bepaalde afwijkingen makkelijker toelaat. Inderdaad worden in MMR gemuteerde tumoren weinig chromosomale afwijkingen aangetroffen. In hoofdstuk 5 bevestigen we de geringe aanwezigheid van chromosomale afwijkingen in een panel van MMR gemuteerde tumor cellijnen in vergelijking met MMR 'wild-type' tumor cellijnen. Om de rol van de genetische achtergrond op het aantal en typen chromosomale afwijkingen te bestuderen introduceren we een modelsysteem waarbij cellijnen met verschillende genetische achtergronden, in dit geval MMR deficiënt versus MMR proficiënt, worden geselecteerd voor methotrexaat (MTX) resistentie. De chromosomale afwijkingen in de resistente cellijnen worden dan geanalyseerd met array CGH. Deze studie laat zien dat de cellijn met een defect MMR systeem meestal geen chromosomale afwijkingen heeft, terwijl de MMR proficiënte cellijn vaak amplificaties vertoont na het selecteren voor MTX resistentie.

Carcinoom van de tuba Fallopii (FTC; eileiderkanker) is een zeldzame en agressieve vorm van kanker. Wij hebben chromosomale afwijkingen bestudeerd in 14 FTC met array CGH. De resultaten van deze studie staan beschreven in hoofdstuk 6. Alle 14 tumoren hebben veel chromosomale afwijkingen, waarbij soms ieder chromosoom van een tumor een afwijking heeft. Een aantal bekende oncogenen liggen in terugkerende amplificaties zoals *CCNE1*, *MYC* en *AKT2*. We merken ook op dat deze tumoren homogeen zijn waarbij een aantal afwijkingen in meer dan 70% van de tumoren voorkomt. Dit suggereert dat er slechts een gelimiteerd aantal routes naar eileiderkanker leidt.

In hoofdstuk 7 beschrijven we de klonale verwantschap tussen de index (primaire) tumor en tweede primaire tumoren in plaveiselcelcarcinomen van de mondholte. Soms ontstaan deze tweede primaire tumoren op een andere anatomische plaats in de mondholte jaren na diagnose van de index tumor. We beschrijven array CGH resultaten van drie patiënten in wie meerdere tumoren in de mondholte ontstonden over een periode van soms meer dan tien jaar. Ondanks de aanwezigheid van tumor specifieke chromosomale afwijkingen was het duidelijk dat tweede primaire tumoren eenzelfde oorsprong hadden gebaseerd op de aanwezigheid van bepaalde chromosomale afwijkingen aanwezig in alle, of een subset van tumoren in iedere patiënt. We beschrijven twee theorieën die deze verwantschap kunnen verklaren. Een voor de hand liggende theorie is dat cellen van de index tumor los komen, zich verspreiden in de mondholte en elders een nieuwe tumor vormen. Ook is het mogelijk dat deze tumoren ontstaan uit een klonale, cellulaire patch waarin alle cellen dezelfde chromosomale afwijking(en) hebben.

Van alle patiënten met plaveiselcelkanker in de mondholte sterft 60% binnen 5 jaar ten gevolge van de ziekte. Deze overlevingskans is in de laatste 40 jaar niet verbeterd. Om meer inzicht te krijgen in de genomische afwijkingen die van invloed zijn op het ontstaan en de ontwikkeling van plaveiselcelkanker hebben we de chromosomale afwijkingen bestudeerd

in 89 tumoren. Deze resultaten staan beschreven in hoofdstuk 8. Tijdens deze studie vonden we negen terugkerende amplificaties kleiner dan 3 Mb. Twee van deze amplificaties zijn frequent aanwezig in plaveiselcelkanker, namelijk, amplificatie van 7p11.2 (*EGFR*) en 11q13 (*CCND1*, *FGFR3*, *FGFR4*, *EMSI* en *PAK1*). Beide amplificaties zijn positief gecorreleerd aan de aanwezigheid van een mutatie in *TP53*. De resterende zeven amplificaties zijn minder frequent aanwezig en niet eerder waargenomen in plaveiselcelkanker van de mondholte. Om te onderzoeken welke moleculaire routes belangrijk zijn voor plaveiselcelkanker hebben we de genexpressie van een aantal genen in deze amplificaties bepaald. Dit onderzoek leidde tot de identificatie van genen betrokken bij integrin signalering (verbinding extracellulaire matrix met cytoskelet), celoverleving, celadhesie en celmigratie, en de Notch en Hedgehog routes. De Notch en Hedgehog routes zijn belangrijk voor patroonvorming tijdens organismale ontwikkeling en celdifferentiatie. Om de rol van deze twee routes in plaveiselcelkanker verder te bestuderen hebben we de genexpressie van een aantal andere genen uit de Notch en Hedgehog routes bepaald. Deze resultaten laten zien dat beide routes frequent gedereguleerd zijn. Deze resultaten zijn mogelijk klinisch relevant, met name voor de ontwikkeling van therapeutische middelen gebaseerd op deregulatie van deze routes in plaveiselcelkanker van de mondholte.

# Bibliography

- Aebi, S., Kurdi-Haidar, B., Gordon, R., Cenni, B., Zheng, H., Fink, D., Christen, R. D., Boland, C. R., Koi, M., Fishel, R., and Howell, S. B. (1996). Loss of dna mismatch repair in acquired resistance to cisplatin. *Cancer Res.*, **56**, 3087–3090.
- Aho, S. (2004). Soluble form of jagged1: unique product of epithelial keratinocytes and a regulator of keratinocyte differentiation. *J Cell Biochem*, **92**, 1271–1281.
- Ai, H., Barrera, J. E., Meyers, A. D., Shroyer, K. R., and Varella-Garcia, M. (2001). Chromosomal aneuploidy precedes morphological changes and supports multifocality in head and neck lesions. *Laryngoscope*, **111**, 1853–1858.
- Akama, Y., Yasui, W., Yokozaki, H., Kuniyasu, H., Kitahara, K., Ishikawa, T., and Tahara, E. (1995). Frequent amplification of the cyclin E gene in human gastric carcinomas. *Jpn J Cancer Res.*, **86**, 617–621.
- Albertson, D. G. (2003). Profiling breast cancer by array cgh. *Breast Cancer Res. Treat*, **78**, 289–298.
- Albertson, D. G., Ylstra, B., SeGRAves, R., Collins, C., Dairkee, S. H., Kowbel, D., Kuo, W.-L., Gray, J. W., and Pinkel, D. (2000). Quantitative mapping of amplicon structure by array cgh identifies cyp24 as a candidate oncogene. *Nat. Genet.*, **25**, 144–146.
- Albertson, D. G., Collins, C., McCormick, F., and Gray, J. W. (2003). Chromosome aberrations in solid tumors. *Nat Genet*, **34**, 369–376.
- Ascano, J., Beverly, L., and Capobianco, A. (2003). The c-terminal pdz-ligand of jagged1 is essential for cellular transformation. *J Biol Chem*, **10**, 8771–8779.
- Aziz, S., Kuperstein, G., Rosen, B., Cole, D., Nedelcu, R., McLaughlin, J., and Narod, S. A. (2001). A genetic epidemiological study of carcinoma of the fallopian tube. *Gynecol. Oncol.*, **80**, 341–345.
- Bailey, J. A., Yavor, A. M., Viggiano, L., Misceo, D., Horvath, J. E., and Archidiacono, N. *et al.* (2002). Human-specific duplication and mosaic transcripts: the recent paralogous structure of chromosome 22. *Am J Hum Genet*, **70**, 83–100.
- Banerjee, D., Mayer-Kuckuk, P., Capiiaux, G., Budak-Alpdogan, T., Gorlick, R., and Bertino, J. R. (2002). Novel aspects of resistance to drugs targeted to dihydrofolate reductase and thymidylate synthase. *Biochim. Biophys. Acta*, **1587**, 164–173.

- Batsakis, J. G. (1984). Synchronous and metachronous carcinomas in patients with head and neck cancer. *Int J Radiat Oncol Biol Phys*, **10**, 2163–2164.
- Beck, J. A., Lloyd, S., Hafezparast, M., Lennon-Pierce, M., Eppig, J. T., Festing, M. F., and Fisher, E. M. (2000). Genomic binding sites of the yeast cell-cycle transcription factors *sbf* and *mbf*. *Nat. Genet.*, **24**, 23–25.
- Bellacosa, A., de Feo, D., Godwin, A. K., Bell, D. W., Cheng, J. Q., Altomare, D. A., Wan, M., Dubeau, L., Scambia, G., and Masciullo, V. *et al.* (1995). Molecular alterations of the *akt2* oncogene in ovarian and breast carcinomas. *Int. J. Cancer*, **64**, 280–285.
- Bocker, T., Ruschoff, J., and Fishel, R. (1999). Molecular diagnostics of cancer predisposition: hereditary non-polyposis colorectal carcinoma and mismatch repair defects. *Biochim. Biophys. Acta*, **1423**, O1–O10.
- Braakhuis, B. J., Tabor, M. P., Leemans, C. R., van der Waal, I., Snow, G. B., and Brakenhoff, R. H. (2002). Second primary tumors and field cancerization in oral and oropharyngeal cancer: molecular techniques provide new insights and definitions. *Head Neck*, **24**, 198–206.
- Bruder, C. E. G., Hirvelä, C., Tapia-Paez, I., Fransson, I., Segraves, R., Hamilton, G., Zhang, X. X., Evans, D. G., Wallace, A. J., Baser, M. E., Zucman-Rossi, J., Hergersberg, M., Bolthausen, E., Papi, L., Rouleau, G. A., Poptodorov, G., Jordanova, A., Rask-Andersen, H. R., Kluwe, L., Mautner, V., Sainio, M., Hung, G., Mathiesen, T., Möller, C., Pulst, S. M., Harder, H., Heiberg, A., Honda, M., Niimura, M., Sahlén, S., Blennow, E., Albertson, D. G., Pinkel, D., and Dumanski, J. P. (2001). High resolution deletion analysis of constitutional dna from neurofibromatosis type 2 (*nf2*) patients using microarray-cgh. *Hum. Mol. Genet.*, **10**, 271–282.
- Buckles, G., Rauskolb, C., Villano, J., and Katz, F. (2001). Four-jointed interacts with *dachs*, *abelson* and *enabled* and feeds back onto the notch pathway to affect growth and segmentation in the drosophila leg. *Development*, **128**, 3533–3542.
- Caduff, R. F., Svoboda-Newman, S. M., Bartos, R. E., Ferguson, A. W., and Frank, T. S. (1998). Comparative analysis of histologic homologues of endometrial and ovarian carcinoma. *Am J Surg Pathol.*, **22**, 319–326.
- Cai, W. W., Reneker, J., Chow, C. W., Vaishnav, M., and Bradley, A. (1998). An anchored framework bac map of mouse chromosome 11 assembled using multiplex oligonucleotide hybridization. *Genomics*, **54**, 387–397.
- Cai, W.-W., Mao, J.-H., Chow, C.-W., Damani, S., Balmain, A., and Bradley, A. (2002). Genome-wide detection of chromosomal imbalances in tumors using bac microarrays. *Nat. Biotech.*, **20**, 393–396.
- Califano, J., Westra, W. H., Meininger, G., Corio, R., Koch, W. M., and Sidransky, D. (2000). Genetic progression and clonal relationship of recurrent premalignant head and neck lesions. *Clin Cancer Res*, **6**, 347–352.
- Carey, T. E. (1996). Field cancerization: are multiple primary cancers monoclonal or polyclonal? *Ann Med*, **28**, 183–188.

- Chaturvedi, V., Chu, M. S., Carrol, B. M., Brenner, B. J., and Nickoloff, B. . (2002). Estimation of size of clonal unit for keratinocytes in normal human skin. *Arch Pathol Lab Med*, **126**, 420–424.
- Chen, Z.-S., Lee, K., S., W., Raftogianis, R. B., Kuwano, M., Zeng, H., and Kruh, G. D. (2002). Analysis of methotrexate and folate transport by multidrug resistance protein 4 (abcc4): Mrp4 is a component of the methotrexate efflux system. *Cancer Res.*, **62**, 3144–3150.
- Cheng, J. Q., Godwin, A. K., Bellacosa, A., Taguchi, T., Franke, T. F., Hamilton, T. C., Tsiichlis, P. N., and Testa, J. R. (1992). Akt2, a putative oncogene encoding a member of a subfamily of protein-serine/threonine kinases, is amplified in human ovarian carcinomas. *Proc. Natl. Acad. Sci. U. S. A.*, **89**, 9267–9271.
- Chernova, O. B., Chernov, M. V., Ishizaka, Y., Agarwal, M. L., and Stark, G. R. (1998). Myc abrogates p53-mediated cell cycle arrest in n-(phosphonacetyl)-l-aspartate-treated cells, permitting cad gene amplification. *Mol. Cell. Biol.*, **18**, 536–545.
- Cheung, J., Wilson, M. D., Zhang, J., Khaja, R., MacDonald, J. R., Heng, H. H., Koop, B. F., and Scherer, S. W. (2003). Recent segmental and gene duplications in the mouse genome. *Genome Biol.*, **4**, R47.
- Cheung, V. G., Nowak, N., Jang, W., Kirsch, I. R., Zhao, S., Chen, X. .-N., Kim, U.-J., Kuo, W.-L., Olivier, M., Conroy, J., Furey, T. S., Kasprzyk, A., Massa, H., Yonescu, R., Sait, S., Thoreen, C., and Snijders, A. *et al.* (2001). Integration of cytogenetic landmarks into the draft sequence of the human genome. *Nature*, **409**, 953–958.
- Chuang, P., Kawcak, T., and McMahon, A. (2003). Feedback control of mammalian hedgehog signaling by the hedgehog-binding protein, hip1, modulates fgf signaling during branching morphogenesis of the lung. *Genes Dev*, **17**, 342–347.
- Chung, T. K., Cheung, T. H., To, K. F., and Wong, Y. F. (2000). Overexpression of p53 and her-2/neu and c-myc in primary fallopian tube carcinoma. *Gynecol. Obstet. Invest.*, **49**, 47–51.
- Cole, P. D., Kamen, B. A., Gorlick, R., Banerjee, D., Smith, A. K., Magill, E., and Bertino, J. (2001). Effects of overexpression of gamma-glutamyl hydrolase on methotrexate metabolism and resistance. *Cancer Res.*, **61**, 4599–4604.
- Cooper, J. S., Pajak, T. F., and Rubin, P. *et al.* (1989). Second malignancies in patients who have head and neck cancer: incidence, effect on survival and implications based on the rtog experience. *Int J Radiat Oncol Biol Phys*, **17**, 449–456.
- Coquelle, A., Pipiras, E., Toledo, F., Buttin, G., and Debatisse, M. (1997). Expression of fragile sites triggers intrachromosomal mammalian gene amplification and sets boundaries to early amplicons. *Cell*, **89**, 215–225.
- Coquelle, A., Toledo, F., Stern, S., Bieth, A., and Debatisse, M. (1998). A new role for hypoxia in tumor progression: induction of fragile site triggering genomic rearrangements and formation of complex dms and hrs. *Mol. Cell*, **2**, 259–265.

- Dev, V. G., Tantravahi, R., Miller, D. A., and Miller, O. J. (1977). Nucleolus organizers in *mus musculus* subspecies and in the rag mouse cell line. *Genetics*, **86**, 389–398.
- Dunham, M. J., Badrane, H., Ferea, T., Adams, J., Brown, P. O., Rosenzweig, F., and Botstein, D. (2002). Characteristic genome rearrangements in experimental evolution of *saccharomyces cerevisiae*. *Proc Natl Acad Sci USA*, **99**, 16144–16149.
- Dutrillaux, B., Muleris, M., and Seureau, M. G. (1986). Imbalance of sex chromosomes, with gain of early-replicating x, in human solid tumors. *Int. J. Cancer*, **38**, 475–479.
- Duval, A. and Hamelin, R. (2002). Mutations at coding repeat sequences in mismatch repair-deficient human cancers: toward a new concept of target genes for instability. *Cancer Res.*, **62**, 2447–2454.
- Eicher, E. M. and Shown, E. P. (1993). Molecular markers that define the distal ends of mouse autosomes 4, 13, and 19 and the sex chromosomes. *Mamm Genome*, **4**, 226–229.
- Eichler, E. E. (2001). Recent duplication, domain accretion and the dynamic mutation of the human genome. *Trends Genet.*, **17**, 661–669.
- Eichler, E. E., Lu, F., Shen, Y., Antonacci, R., Jurecic, V., and Doggett, N. A. *et al.* (1996). Duplication of a gene-rich cluster between 16p11.1 and xq28: a novel pericentromeric-directed mechanism for paralogous genome evolution. *Hum Mol Genet.*, **5**, 899–912.
- Elsevier, S. M. and Ruddle, F. H. (1975). Location of genes coding for 18s and 28s ribosomal rna within the genome of *mus musculus*. *Chromosoma*, **52**, 219–228.
- Emanuel, B. S. and Shaikh, T. H. (2001). Segmental duplications: an 'expanding' role in genomic instability and disease. *Nat Rev Genet.*, **2**, 791–800.
- Eshleman, J. R., Casey, G., Kochera, M. E., Sedwick, W. D., Swinler, S. E., Veigl, M. L., Willson, J. K. V., Schwartz, S., and Markowitz, S. D. (1998). Chromosome number and structure both are markedly stable in rer colorectal cancers and are not destabilized by mutation of p53. *Oncogene*, **17**, 719–725.
- Esteller, M. (2000). Epigenetic lesions causing genetic lesions in human cancer: promoter hypermethylation of dna repair genes. *Eur. J. Cancer*, **36**, 2294–2300.
- Ewart-Toland, A., Briassouli, P., de Koning, J. P., Mao, J. H., Yuan, J., and Chan, F. *et al.* (2003). Identification of *stk6/stk15* as a candidate low-penetrance tumor-susceptibility gene in mouse and human. *Nat Genet.*, **34**, 403–412.
- Fiegler, H., Philpa, C., Douglas, E. J., Burford, D. C., Hunt, S., Smith, J., Vetrie, D., Gorman, P., Tomlinson, I. P. M., and Carter, N. P. (2003). Dna microarrays for comparative genomic hybridization based on dop-pcr amplification of bac and pac clones. *Genes Chromosomes Cancer*, **36**, 361–374.
- Fink, D., Nebe, I. S., Norris, P. S., Baergen, R. N., Wilczynsk, S. P., Costa, M. J., Haas, M., Cannistra, S. A., and Howell, S. B. (1998). Enrichment for dna mismatch repair-deficient cells during treatment with cisplatin. *Int. J. Cancer*, **77**, 741–746.
- Fridlyand, J., Snijders, A. M., Pinkel, D., Albertson, D. G., and Jain, A. N. (2004). Hidden markov models approach to the analysis of array cgh data. *J Multivariate Anal.*, **in press**.



- Frouin, I., Prosperi, E., Denegri, M., Negri, C., Donzelli, M., Rossi, L., Riva, F., Stefanini, M., and Scovassi, A. I. (2001). Different effects of methotrexate on dna mismatch repair proficient and deficient cells. *Eur. J. Cancer*, **37**, 1173–1180.
- Gavish, D., Azrolan, N., and Breslow, J. L. (1989). Plasma ip(a) concentration is inversely correlated with the ratio of kringle iv/kringle v encoding domains in the apo(a) gene. *J Clin Invest.*, **84**, 2021–2027.
- Geisen, C. and Möröy, T. (2002). The oncogenic activity of cyclin e is not confined to cdk2 activation alone but relies on several other, distinct functions of the protein. *J. Biol. Chem.*, **277**, 39909–39918.
- Geschwind, D. H., Gregg, J., Boone, K., Karrim, J., Pawlikowska-Haddal, A., Rao, E., Ellison, J., Ciccodicola, A., D’Urso, M., Woods, R., Rappold, G. A., Swerdloff, R., and Nelson, S. F. (1998). Klinefelter’s syndrome as a model of anomalous cerebral laterality: testing gene dosage in the x chromosome pseudoautosomal region using a dna microarray. *Dev. Genet.*, **23**, 215–229.
- Gray, G., Mann, R., Mitsiadis, E., Henrique, D., Carcangiu, M., and Banks, A. (1999). Human ligands of the notch receptor. *Am J Pathol*, **154**, 785–794.
- Guan, X., Xu, J., Anzick, S., Zhang, H., Trent, J. M., and Meltzer, P. S. (1999). Hybrid selection of transcribed sequences from microdissected dna: isolation of genes within amplified region at 20q11-q13.2 in breast cancer. *Cancer Res.*, **56**, 3446–3450.
- Guenet, J. L. and Bonhomme, F. (2003). Wild mice: an ever-increasing contribution to a popular mammalian model. *Trends Genet*, **19**, 24–31.
- Haas, K., Johannes, C., Geisen, C., Schmidt, T., Karsunky, H., Blass-Kampmann, S., Obe, G., and Möröy, T. (1997). Malignant transformation by cyclin e and ha-ras correlates with lower sensitivity towards induction of cell death but requires functional myc and cdk4. *Oncogene*, **15**, 2615–2623.
- Hanahan, D. and Weinberg, R. (2000). The hallmarks of cancer. *Cell*, **100**, 57–70.
- Hawn, M. T., Umar, A., Carethers, J. M., Marra, G., Kunkel, T. A., Boland, C. R., and Koi, M. (1995). Evidence for a connection between the mismatch repair system and the g2 cell cycle checkpoint. *Cancer Res.*, **55**, 3721–3725.
- Heiskanen, M. A., Bittner, M. L., Chen, Y., Khan, J., Adler, K. E., Trent, J. M., and Meltzer, P. S. (2000). Detection of gene amplification by genomic hybridization to cdna microarrays. *Cancer Res.*, **60**, 799–802.
- Hellman, A., Zlotorynski, E., Scherer, S. W., Cheung, J., Vincent, J. B., Smith, D. I., Trakhtenbrot, L., and Kerem, B. (2002). A role for common fragile site induction in amplification of human oncogenes. *Cancer Cell*, **1**, 89–97.
- Hellström, A. C., Hue, J., Silfverswärd, C., and Auer, G. (1994). Dna-ploidy and mutant p53 overexpression in primary fallopian tube cancer. *Int. J. Gynecol. Cancer*, **4**, 408–413.
- Henderson, A. S., Eicher, E. M., Yu, M. T., and Atwood, K. C. (1974). The chromosomal location of ribosomal dna in the mouse. *Chromosoma*, **49**, 155–160.

- Heselmeyer, K., Hellström, A.-C., Blegen, H., Schröck, E., Silfverswärd, C., Shah, K., Auer, G., and Ried, T. (1998). Primary carcinoma of the fallopian tube: comparative genomic hybridization reveals high genetic instability and a specific, recurring pattern of chromosomal aberrations. *Int. J. Gynecol. Pathol.*, **17**, 245–254.
- Hodgson, G., Hager, J. H., Volik, S., Hariono, S., Wernick, M., Moore, D., Albertson, D. G., Pinkel, D., Collins, C., Hanahan, D., and Gray, J. W. (2001). Genome scanning with array cgh delineates regional alterations in mouse islet carcinomas. *Nat. Genet.*, **29**, 459–464.
- Hollox, E., Armour, A., and Barber, C. (2003). Extensive normal copy number variation of a  $\beta$ -defensin antimicrobial-gene cluster. *Am. J. Hum. Genet.*, **73**, 591–600.
- Horak, C. E., Mahajan, M. C., Luscombe, N. M., Gerstein, M., Weissman, S. M., and Snyder, M. (2002). Gata-1 binding sites mapped in the  $\gamma$ -globin locus by using mammalian chip-chip analysis. *Proc. Natl. Acad. Sci. USA*, **99**, 2924–2929.
- Horvath, J. E., Bailey, J. A., Locke, D. P., and Eichler, E. E. (2001). Lessons from the human genome: transitions between euchromatin and heterochromatin. *Hum Mol Genet.*, **10**, 2215–2223.
- Huang, Q., Yu, G., McCormick, S., Mo, J., Datta, B., and Mahimkar, M. *et al.* (2002). Genetic differences detected by comparative genomic hybridization in head and neck squamous cell carcinomas from different tumor sites: construction of oncogenetic trees for tumor progression. *Genes Chromosomes Cancer*, **34**, 224–233.
- Hume, W. and Potten, C. (1979). Advances in epithelial kinetics—an oral view. *J Oral Pathol*, **8**, 3–22.
- Hyman, E., Kauraniemi, P., Hautaniemi, S., Wolf, M., Mousses, S., Rozenblum, E., Ringnér, M., Sauter, G., Monni, O., Elkahloun, A., Kallioniemi, O. P., and Kallioniemi, A. (2002). Impact of dna amplification on gene expression patterns in breast cancer. *Cancer Res.*, **62**, 6240–6245.
- Infante, J. J., Dombek, K. M., Rebordinos, L., Cantoral, J. M., and Young, E. T. (2003). Genome-wide amplifications caused by chromosomal rearrangements play a major role in the adaptive evolution of natural yeast. *Genetics*, **165**, 1745–1759.
- Innocente, S. A., Abrahamson, J. L. A., Cogswell, J. P., and Lee, J. M. (1999). p53 regulates a g2 checkpoint through cyclin b1. *Proc. Natl. Acad. Sci. USA*, **96**, 2147–2152.
- Iyer, V. R., Horak, C. E., Scafe, C. S., Botstein, D., Snyder, M., and Brown, P. O. (2001). Genomic binding sites of the yeast cell-cycle transcription factors sbf and mbf. *Nature*, **409**, 533–538.
- Jackson, M. S., Rocchi, M., Thompson, G., Hearn, T., Crosier, M., and Guy, J. *et al.* (1999). Sequences flanking the centromere of human chromosome 10 are a complex patchwork of arm-specific sequences, stable duplications and unstable sequences with homologies to telomeric and other centromeric locations. *Hum Mol Genet.*, **8**, 205–215.
- Jain, A. N., Tokuyasu, T., Snijders, A. M., Segraves, R., Albertson, D. G., and Pinkel, D. (2001). Fully automatic quantification of microarray image data. *Genome Res.*, **12**, 325–332.

- Jang, S. J., Chiba, I., Hirai, A., Hong, W. K., and Mao, L. (2001). Multiple oral squamous epithelial lesions: are they genetically related? *Oncogene*, **20**, 2235–2242.
- Ji, Y., Eichler, E. E., Schwartz, S., and Nicholls, R. D. (2000). Structure of chromosomal duplicons and their role in mediating human genomic disorders. *Genome Res.*, **10**, 597–610.
- Jonason, A. S., Kunala, S., and Price, G. J. *et al.* (1996). Frequent clones of p53-mutated keratinocytes in normal human skin. *Proc Natl Acad Sci U S A*, **93**, 14025–14029.
- Kallioniemi, A., Kallioniemi, O.-P., Sudar, D., Rutovitz, D., Gray, J. W., Waldman, F., and Pinkel, D. (1992). Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science*, **258**, 818–821.
- Kallioniemi, O. P., Kallioniemi, A., Piper, J., Isola, J., Waldman, F. M., Gray, J. W., and Pinkel, D. (1994). Optimizing comparative genomic hybridization for analysis of dna sequence copy number changes in solid tumors. *Genes Chromosomes Cancer*, **10**, 231–243.
- Karhu, R., Kähkönen, M., Kuukasjärvi, T., Pennanen, S., Tirkkonen, M., and Kallioniemi, O. (1997). Quality control of cgh: impact of metaphase chromosomes and the dynamic range of hybridization. *Cytometry*, **28**, 198–205.
- Kauraniemi, P., Bärnlund, M., Monni, O., and Kallioniemi, A. (2001). New amplified and highly expressed genes discovered in the erbb2 amplicon in breast cancer by cdna microarrays. *Cancer Res.*, **61**, 8235–8240.
- Kelly, R., Gibbs, M., Collick, A., and Jeffreys, A. J. (1991). Spontaneous mutation at the hypervariable mouse minisatellite locus ms6-hm: flanking dna sequence and analysis of germline and early somatic mutation events. *Proc R Soc Lond B Biol Sci.*, **245**, 235–245.
- Kipling, D., Salido, E. C., Shapiro, L. J., and Cooke, H. J. (1996). High frequency de novo alterations in the long-range genomic structure of the mouse pseudoautosomal region. *Nat Genet.*, **13**, 78–80.
- Kirchhoff, M., Gerdes, T., Maahr, J., Rose, H., Bentz, M., Dohner, H., and Lundsteen, C. (1999). Deletions below 10 megabasepairs are detected in comparative genomic hybridization by standard reference intervals. *Genes Chromosomes Cancer*, **25**, 410–413.
- Kitahara, K., Yasui, W., Kuniyasu, H., Yokozaki, H., Akama, Y., Yunotani, S., Hisatsugu, T., and E., T. (1995). Concurrent amplification of cyclin e and cdk2 genes in colorectal carcinomas. *Int. J. Cancer*, **62**, 25–28.
- Klein, C. A., Schmidt-Kittler, O., Schardt, J. A., Pantel, K., Speicher, M. R., and Riethmuller, G. (1999). Comparative genomic hybridization, loss of heterozygosity, and dna sequence analysis of single cells. *Proc. Natl. Acad. Sci. USA*, **96**, 4494–4499.
- Klein, O. D., Cotter, P. D., Albertson, D. G., Pinkel, D., Tidyman, W. E., Moore, M. W., and Rauen, K. A. (2004). Prader-willi syndrome resulting from an unbalanced translocation: characterization by array comparative genomic hybridization. *Clin Genet.*, **65**, 477–482.

- Knight, S. J., Lese, C. M., Precht, K. S., Kuc, J., Ning, Y., Lucas, S., Regan, R., Brenan, M., Nicod, A., Lawrie, N. M., Cardy, D. L., Nguyen, H., Hudson, T. J., Riethman, H. C., Ledbetter, D. H., and Flint, J. (2000). An optimized set of human telomere clones for studying telomere integrity and architecture. *Am. J. Hum. Genet.*, **67**, 320–332.
- Koi, M., Umar, A., Chauhan, D. P., Cherian, S. P., Carethers, J. M., Kunkel, T. A., and Boland, C. R. (1994). Human chromosome 3 corrects mismatch repair deficiency and microsatellite instability and reduces n-methyl-n'-nitro-n-nitrosoguanidine tolerance in colon tumor cells with homozygous hmlh1 mutation. *Cancer Res.*, **54**, 4308–4312.
- Koide, T., Ishiura, M., Hazumi, N., Shiroishi, T., Okada, Y., and Uchida, T. (1990). Amplification of a long sequence that includes a processed pseudogene for elongation factor 2 in the mouse. *Genomics*, **6**, 80–88.
- Koide, T., Yoshino, M., Niwa, M., Ishiura, M., Shiroishi, T., and Moriwaki, K. (1992). The amplified long genomic sequence (algs) located in the centromeric regions of mouse chromosomes. *Genomics*, **13**, 1186–1191.
- Koolen, D. A., Vissers, L. E., Nillesen, W., Smeets, D., van Ravenswaaij, C. M., Sistermans, E. A., Veltman, J. A., and de Vries, B. D. (2004). A novel microdeletion, del(2)(q22.3q23.3) in a mentally retarded patient, detected by array-based comparative genomic hybridization. *Clin Genet.*, **65**, 429–432.
- Krause, K., Wasner, M., Reinhard, W., Haugwitz, U., Dohna, C. L., Mssner, J., and Engeland, K. (2000). The tumour suppressor protein p53 can repress transcription of cyclin b. *Nucleic Acids Res.*, **28**, 4410–4118.
- Kuo, M. T., Vyas, R. C., Jiang, L.-X., and Hittelman, W. N. (1994). Chromosome breakage at a major fragile site associated with p-glycoprotein gene amplification in multidrug-resistant cho cells. *Mol. Cell. Biol.*, **14**, 5202–5211.
- Kurihara, Y., Suh, D. S., Suzuki, H., and Moriwaki, K. (1994). Chromosomal locations of ag-nors and clusters of ribosomal dna in laboratory strains of mice. *Mamm Genome*, **5**, 225–228.
- Lander, E. S. *et al.* (2001). The international human genome sequence consortium: Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
- Leeper, K., Garcia, R., Swisher, E., Goff, B., Greer, B., and Paley, P. (2002). Pathologic findings in prophylactic oophorectomy specimens in high-risk women. *Gynecol. Oncol.*, **87**, 52–56.
- Leethanakul, C., Patel, V., Gillespie, J., Pallente, M., Ensley, J., and Koontongkaew, S. *et al.* (2000). Distinct pattern of expression of differentiation and growth-related genes in squamous cell carcinomas of the head and neck revealed by the use of laser capture microdissection and cDNA arrays. *Oncogene*, **19**, 3220–3224.
- Lefort, K. and Dotto, G. (2004). Notch signaling in the integrated control of keratinocyte growth/differentiation and tumor suppression. *Semin Cancer Biol*, **14**, 374–386.

- Lin, L., Prescott, M. S., Zhu, Z., Singh, P., Chun, S. Y., Kuick, R. D., Hanash, S. M., Orringer, M. B., Glover, T. W., and Beer, D. G. (2000). Identification and characterization of a 19q12 amplicon in esophageal adenocarcinomas reveals cyclin E as the best candidate gene for this amplicon. *Cancer Res.*, **60**, 7021–7027.
- Locke, D. P., Segraves, R., Carbone, L., Archidiacono, N., Albertson, D. G., Pinkel, D., and Eichler, E. E. (2003). Large-scale variation among human and great ape genomes determined by array comparative genomic hybridization. *Genome Res.*, **13**, 347–357.
- Locke, D. P., Segraves, R., Nicholls, R. D., Schwartz, S., Pinkel, D., Albertson, D. G., and Eichler, E. E. (2004). Bac microarray analysis of 15q11-q13 rearrangements and the impact of segmental duplications. *J Med Genet.*, **41**, 175–182.
- Loeb, L. A. (2001). A mutator phenotype in cancer. *Cancer Res.*, **61**, 3230–3239.
- Lohi, J. (2001). Laminin-5 in the progression of carcinomas. *Int J Cancer*, **94**, 763–767.
- Lucito, R. and Wigler, M. (2002). *DNA microarrays: a molecular cloning manual*. Cold Spring Harbor Laboratory Press. D. Bowtell and J. Sambrook, eds., 386–393.
- Lucito, R., West, J., Reiner, A., Alexander, J., Esposito, D., Mishra, B., Powers, S., Norton, L., and Wigler, M. (2000). Detecting gene copy number fluctuations in tumor cells by microarray analysis of genomic representations. *Genome Res.*, **10**, 1726–1736.
- Lupski, J. R. (1998). Genomic disorders: structural features of the genome can lead to dna rearrangements and human disease traits. *Trends Genet.*, **14**, 417–422.
- Manavi, M., Berger, A., Kucera, E., Schneeweiss, A., Kucera, H., Kubista, E., and Czerwenka, K. (1998). Amplification and expression of the c-erbB-2 oncogene in mullerian-derived genital-tract tumors. *Gynecol. Oncol.*, **71**, 165–171.
- Manuelidis, L. (1981). Consensus sequence of mouse satellite dna indicates it is derived from tandem 116 basepair repeats. *FEBS Lett.*, **129**, 25–28.
- Marone, M., Scambia, G., Giannitelli, C., Ferrandina, G., Masciullo, V., Bellacosa, A., Benedetti-Panici, P., and Mancuso, S. (1998). Analysis of cyclin e and cdk2 in ovarian cancer: gene amplification and rna overexpression. *Int J Cancer.*, **75**, 34–39.
- Matherly, L. (2001). Molecular and cellular biology of the human reduced folate carrier. *Prog. Nucl. Acid Res.*, **67**, 131–162.
- Matsuda, Y. and Chapman, V. M. (1991). In situ analysis of centromeric satellite dna segregating in mus species crosses. *Mamm Genome.*, **1**, 71–77.
- Mineta, H., Borg, A., Dictor, M., Wahlberg, P., and Wennerberg, J. (1997). Correlation between p53 mutation and cyclin d1 amplification in head and neck squamous cell carcinoma. *Oral Oncol.*, **33**, 42–46.
- Mizuuchi, M., Mori, Y., Sato, K., Kamiya, H., Okamura, N., Nasim, S., Garrett, C. T., and Kudo, R. (1995). High incidence of point mutation in k-ras codon 12 in carcinoma of the fallopian tube. *Cancer*, **76**, 86–90.

- Monk, B. J., Chapman, J. A., Johnson, G. A., Brightman, B. K., Wilczynski, S. P., Schell, M. J., and Fan, H. (1994). Correlation of c-myc and her-2/neu amplification and expression with histopathologic variables in uterine corpus cancer. *Am. J. Obstet. Gynecol.*, **171**, 1193–1198.
- Muleris, M., Dutrillaux, A. M., Olschwang, S., Salmon, R. J., and Dutrillaux, B. (1995). Predominance of normal karyotype in colorectal tumors from hereditary non-polyposis colorectal cancer patients. *Genes Chromosomes Cancer*, **14**, 223–226.
- Muleris, M. and Salmon, R. J. and Dutrillaux, B. (1990). Cytogenetics of colorectal adenocarcinomas. *Cancer Genet. Cytogenet.*, **46**, 143–156.
- Nagase, H., Bryson, S., Cordell, H., Kemp, C. J., Fee, F., and Balmain, A. (1995). Distinct genetic loci control development of benign and malignant skin tumours in mice. *Nat Genet*, **10**, 424–429.
- Nayal, A., Webb, D., and Horwitz, A. (2004). Talin: an emerging focal point of adhesion dynamics. *Curr Opin Cell Biol*, **16**, 94–98.
- Nicolas, M., Wolfer, A., Raj, K., Kummer, J., Mill, P., and van Noort, M. *et al.* (2003). Notch1 functions as a tumor suppressor in mouse skin. *Nat Genet*, **33**, 416–421.
- Nishimaki, H., Kasai, K., Kozaki, K., Takeo, T., Ikeda, H., and Saga, S. *et al.* (2004). A role of activated sonic hedgehog signaling for the cellular proliferation of oral squamous cell carcinoma cell line. *Biochem Biophys Res Commun*, **314**, 313–320.
- Nyberg, K. A., Michelson, R. J., Putnam, C. W., and Weinert, T. A. (2002). Toward maintaining the genome: Dna damage and replication checkpoints. *Annu Rev Genet.*, **36**, 617–656.
- Ohki, K., Kumamoto, H., Ichinohasama, R., Sato, T., Takahashi, N., and Ooya, K. (2004). Ptc gene mutations and expression of shh, ptc, smo, and gli-1 in odontogenic keratocysts. *Int J Oral Maxillofac Surg*, **33**, 584–592.
- Osoegawa, K., Tateno, M., Woon, P. Y., Frengen, E., Mammoser, A. G., and Catanese, J. J. *et al.* (2000). Bacterial artificial chromosome libraries for mouse sequencing and functional analysis. *Genome Res.*, **10**, 116–128.
- Owens, D. M. and Watt, F. M. (2003). Contribution of stem cells and differentiated cells to epidermal tumours. *Nat Rev Cancer*, **3**, 444–451.
- Pardue, M. L. and Gall, J. G. (1970). Chromosomal localization of mouse satellite dna. *Science*, **168**, 1356–1358.
- Parkin, D., Pisani, P., and Ferlay, J. (1999). Global cancer statistics. *CA Cancer J Clin*, **49**, 33–64.
- Partridge, M., Pateromichelakis, S., Phillips, E., Emilion, G., and Langdon, J. (2001). Profiling clonality and progression in multiple premalignant and malignant oral lesions identifies a subgroup of cases with a distinct presentation of squamous cell carcinoma. *Clin Cancer Res*, **7**, 1860–1866.

- Paulson, T. G., Almasan, A., Brody, L. L., and Wahl, G. M. (1998). Gene amplification in a p53-deficient cell line requires cell cycle progression under conditions that generate dna breakage. *Mol. Cell. Biol.*, **18**, 3089–3100.
- Pelengaris, S., Khan, M., and Evan, G. I. (2002). Suppression of myc-induced apoptosis in beta cells exposes multiple oncogenic properties of myc and triggers carcinogenic progression. *Cell*, **109**, 321–334.
- Pere, J., Tapper, J., Seppälä, M., Knuutila, S., and Butzow, R. (1998). Genomic alterations in fallopian tube carcinoma: comparison to serous uterine and ovarian carcinomas reveals similarity suggesting likeness in molecular pathogenesis. *Cancer Res.*, **58**, 4274–4276.
- Piek, J. M. M., van Diest, P. J., Zweemer, R. P., Jansen, J. W., Poort-Keesom, R. J. J., Menko, F. H., Gille, J. J. P., Jongsma, A. P. M., Pals, G., Kenemans, P., and Verheijen, R. H. M. (2001). Dysplastic changes in prophylactically removed fallopian tubes of women predisposed to developing ovarian cancer. *J. Pathol.*, **195**, 451–456.
- Pinkel, D., Segraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W.-L., Chen, C., Zhai, Y., Dairkee, S. H., Ljung, B.-M., Gray, J. W., and Albertson, D. G. (1998). High resolution analysis of dna copy number variation using comparative genomic hybridization to microarrays. *Nature Genet.*, **20**, 207–211.
- Pipiras, E., Coquelle, A., Bieth, A., and Debatisse, M. (1998). Interstitial deletions and intrachromosomal amplification initiated from a double-strand break targeted to a mammalian chromosome. *EMBO J.*, **17**, 325–333.
- Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D., and Brown, P. O. (1999). Genome-wide analysis of dna copy-number changes using cdna microarrays. *Nature Genet.*, **23**, 41–46.
- Rabiner, L. R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, **77**, 257–286.
- Rauen, K. A., Albertson, D. G., Pinkel, D., and Cotter, P. D. (2002). Additional patient with del(12)(q21.2q22): Further evidence for a candidate region for cardio-facio-cutaneous syndrome? *Am. J. Med. Genet.*, **110**, 51–56.
- Remvikos, Y., Vogt, N., Muleris, M., Salmon, R. J., Malfoy, B., and Dutrillaux, B. (1995). Dna-repeat instability is associated with colorectal cancers presenting minimal chromosome rearrangements. *Genes Chromosomes Cancer*, **12**, 272–276.
- Richter, J., Wagner, U., Kononen, J., Fijan, A., Bruderer, J., Schmid, U., Ackermann, D., Maurer, R., Alund, G., Knonagel, H., Rist, M., Wilber, K., Anabitarte, M., Hering, F., Hardmeier, T., Schonenberger, A., Flury, R., Jager, P., Fehr, J. L., Schraml, P., Moch, H., Mihatsch, M. J., Gasser, T., Kallioniemi, O. P., and Sauter, G. (2000). High-throughput tissue microarray analysis of cyclin e gene amplification and overexpression in urinary bladder cancer. *Am. J. Pathol.*, **157**, 787–794.
- Rots, M. G., Pieters, R., Peters, G. J., Noordhuis, P., van Zantwijk, C. H., Kaspers, G. J. L., Hahlen, K., Creutzig, U., Veerman, A. J. P., and Jansen, G. (1999). Circumvention of methotrexate resistance in childhood leukemia subtypes by rationally designed antifolates. *Blood*, **93**, 1677–1683.

- Ruiz i Altaba, A., Sanchez, P., and Dahmane, N. (2002). Gli and hedgehog in cancer: tumours, embryos and stem cells. *Nat Rev Cancer*, **2**, 361–372.
- Samonte, R. V. and Eichler, E. E. (2002). Segmental duplications and the evolution of the primate genome. *Nat Rev Genet.*, **3**, 65–72.
- Schaeffer, A. J., Chung, J., Heretis, K., Wong, A., Ledbetter, D. H., and Lese Martin, C. (2004). Comparative genomic hybridization-array analysis enhances the detection of aneuploidies and submicroscopic imbalances in spontaneous miscarriages. *Am J Hum Genet.*, **74**, 1168–1174.
- Schlegel, J., Stumm, G., Scherthan, H., Bocker, T., Zirngibl, H. R., Rschoff, J., and Hofstädter, F. (1995). Comparative genomic in situ hybridization of colon carcinomas with replication error. *Cancer Res.*, **55**, 6002–6005.
- Shaw, C. J. and Lupski, J. R. (2004). Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. *Hum Mol Genet.*, **13 Suppl 1**, R57–64.
- Shi, H., Yan, P. S., Chen, C. M., Rahmatpanah, F., Lofton-Day, C., Caldwell, C. W., and Hui-Ming Huang, T. (2002). Expressed cpg island sequence tag microarray for dual screening of dna hypermethylation and gene silencing in cancer cells. *Cancer Res.*, **62**, 3214–3220.
- Sidman, C. L. and Shaffer, D. J. (1994). Large-scale genomic comparison using two-dimensional dna gels. *Genomics*, **23**, 15–22.
- Singer, M. J., Mesner, L. D., Friedman, C. L., Trask, B. J., and Hamlin, J. L. (2000). Amplification of the human dihydrofolate reductase gene via double minutes is initiated by chromosome breaks. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 7921–7926.
- Slaughter, D. P., Southwick, H. W., and Smejkal, W. (1953). "field cancerization" in oral stratified squamous epithelium: clinical implications of multicenter origin. *Cancer*, **6**, 963.
- Snedecor, G. W. and Cochran, W. G. (1989). *Statistical Methods*. Iowa State University Press.
- Snijders, A. M., Nowak, N., Segraves, R., Blackwood, S., Brown, N., Conroy, J., Hamilton, G., Hindle, A. K., Huey, B., Kimura, K., Law, S., Myambo, K., Palmer, J., Ylstra, B., Yue, J. P., Gray, J. W., Jain, A. N., Pinkel, D., and Albertson, D. G. (2001). Assembly of microarrays for genome-wide measurement of dna copy number. *Nature Genet.*, **29**, 263–264.
- Snijders, A. M., Fridlyand, J., Mans, D. A., Segraves, R., Jain, A. N., Pinkel, D., and G., A. D. (2003). Shaping of tumor and drug resistant genomes by instability and selection. *Oncogene*, **22**, 4370–4379.
- Solinas-Toldo, S., Lampel, S., Stilgenbauer, S., Nickolenko, J., Benner, A., Dohner, H., Cremer, T., and Lichter, P. (1997). Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosomes Cancer*, **20**, 399–407.



- Soulie, P., Fourme, E., Hamelin, R., Asselain, B., Salmon, R.-J., Dutrillaux, B., and Muleris, M. (1999). Tp53 status and gene amplification in human colorectal carcinomas. *Cancer Genet. Cytogenet.*, **115**, 118–122.
- Spiteri, E., Babcock, M., Kashork, C. D., Wakui, K., Gogineni, S., and Lewis, D. A. *et al.* (2003). Frequent translocations occur between low copy repeats on chromosome 22q11.2 (1cr22s) and telomeric bands of partner chromosomes. *Hum Mol Genet.*, **12**, 1823–1837.
- Spruck, C. H., Won, K.-A., and Reed, S. (1999). Deregulated cyclin e induces chromosome instability. *Nature*, **401**, 297–300.
- Squier, C. A. and Kremer, M. J. (2001). Biology of oral mucosa and esophagus. *J Natl Cancer Inst Monogr*, **29**, 7–15.
- Srimatkandada, S., Dube, S. K., Carmen, M., and R., B. J. (2000). Coamplification of 3-hydroxy-3-methylglutaryl coenzyme a reductase genes in methotrexate-resistant human leukemia cell lines. *Oncol. Res.*, **12**, 11–15.
- Staelens, J., Wielockx, B., Puimege, L., van Roy, F., Guenet, J. L., and Libert, C. (2002). Hyporesponsiveness of sprete mice to lethal shock induced by tumor necrosis factor and implications for a tnf based antitumor therapy. *Proc Natl Acad Sci U S A*, **99**, 3940–3945.
- Sturgis, E. M. and Miller, R. H. (1995). Second primary malignancies in the head and neck cancer patient. *Ann Otol Rhinol Laryngol*, **104**, 946–954.
- Sudbo, J., Lippman, S. M., and Lee, J. J. *et al.* (2004). The influence of resection and aneuploidy on mortality in oral leukoplakia. *N Engl J Med*, **350**, 1405–1413.
- Sun, P. C., Uppaluri, R., Schmidt, A. P., Pashia, M. E., Quant, E. C., Sunwoo, J. B., Gollin, S. M., and Scholnick, S. B. (2001). Transcript map of the 8p23 putative tumor suppressor region. *Genomics*, **75**, 17–25.
- Sutherland, G. R. (1991). Chromosomal fragile sites. *GATA*, **8**, 161–166.
- Sutherland, G. R. and Richards, R. I. (1995). The molecular basis of fragile sites in human chromosomes. *Curr. Opin. Genet. Dev.*, **5**, 323–327.
- Sutherland, G. R., Baker, E., and Richards, R. I. (1998). Fragile sites still breaking. *Trends Genet.*, **14**, 501–506.
- Tabor, M. P., Brakenhoff, R. H., and Ruijter-Schippers, H. J. *et al.* (2002). Multiple head and neck tumors frequently originate from a single preneoplastic lesion. *Am J Pathol*, **161**, 1051–1060.
- Takemura, Y., Kobayashi, H., and Miyachi, H. (1999). Variable expression of the folylpolyglutamate synthetase gene at the level of mrna transcription in human leukemia cell lines sensitive, or made resistant, to various antifolate drugs. *Anti-Cancer Drugs*, **10**, 677–683.
- Taylor, W. R. and Stark, G. R. (2001). Regulation of the g2/m transition by p53. *Oncogene*, **20**, 1803–1815.

- Telenius, H., Carter, N. P., Bebb, C. E., Nordenskjold, M., Ponder, B. A., and Tunnacliffe, A. (1992). Degenerate oligonucleotide-primed pcr: general amplification of target dna by a single degenerate primer. *Genomics*, **13**, 718–725.
- Thomas, J. W., Schueler, M. G., Summers, T. J., Blakesley, R. W., McDowell, J. C., and Thomas, P. J. *et al.* (2003). Pericentromeric duplications in the laboratory mouse. *Genome Res.*, **13**, 55–63.
- Turner, F. B., Taylor, S. M., and Moran, R. G. (2000). Expression patterns of the multiple transcripts from the foylpolylglutamate synthetase gene in human leukemias and normal differentiated tissues. *J. Biol. Chem.*, **275**, 35960–35968.
- Tuzun, E., Bailey, J. A., and Eichler, E. E. (2004). Recent segmental duplications in the working draft assembly of the brown norway rat. *Genome Res.*, **14**, 493–506.
- Veeraraghavalu, K., Pett, M., Kumar, R., Nair, P., Rangarajan, A., Stanley, M., and Krishna, S. (2004). Papillomavirus-mediated neoplastic progression is associated with reciprocal changes in jagged1 and manic fringe expression linked to notch activation. *J Virol*, **78**, 8687–8700.
- Veltman, J. A., Schoenmakers, E. F. P. M., Eussen, B. H., Janssen, I., Merkx, G., van Cleef, B., van Ravenswaaij, C. M., Brunner, H. G., Smeets, D., and van Kessel, A. D. (2002). High-throughput analysis of subtelomeric chromosome rearrangements by use of array-based comparative genomic hybridization. *Am. J. Hum. Genet.*, **70**, 1269–1276.
- Vissers, L. E., de Vries, B. B., Osoegawa, K., Janssen, I. M., Feuth, T., Choy, C. O., Straatman, H., and van der Vliet, W. *et al.* (2003). Array-based comparative genomic hybridization for the genomewide detection of submicroscopic chromosomal abnormalities. *Am J Hum Genet.*, **73**, 1261–1270.
- Wade, C. M., Kulbokas, E. J. r., Kirby, A. W., Zody, M. C., Mullikin, J. C., Lander, E. S., Lindblad-Toh, K., and Daly, M. J. (2002). The mosaic structure of variation in the laboratory mouse genome. *Nature*, **420**, 574–578.
- Waldman, F. M., DeVries, S., Chew, K. L., Moore, D. H. n., Kerlikowske, K., and Ljung, B. M. (2000). Chromosomal alterations in ductal carcinomas in situ and their in situ recurrences. *J. Natl. Cancer Inst.*, **92**, 313–320.
- Wang, Z. R., Liu, W., Smith, S. T., Parrish, R. S., and Young, S. R. (1999). c-myc and chromosome 8 centromere studies of ovarian cancer by interphase fish. *Exp. Mol. Pathol.*, **66**, 140–148.
- Waridel, F., Estreicher, A., and Bron, L. *et al.* (1997). Field cancerisation and polyclonal p53 mutation in the upper aero-digestive tract. *Oncogene*, **14**, 163–169.
- Watkins, D. and Peacock, C. (2004). Hedgehog signalling in foregut malignancy. *Biochem Pharmacol*, **68**, 1055–1060.
- Weiss, M. M., Hermsen, M. A. J. A., Meijer, G. A., van Grieken, N. C. T., Baak, J. P. A., J., K. E., and van Diest, P. J. (1999). Comparative genomic hybridisation. *J. Clin. Pathol. - Mol. Pathol.*, **52**, 243–251.

- Weng, A. and Aster, J. (2004). Multiple niches for notch in cancer: context is everything. *Curr Opin Genet Dev*, **14**, 48–54.
- Wessendorf, S., Fritz, B., Wrobel, G., Nessling, M., Lampel, S., Göettel, D., Kùepper, M., Joos, S., Hopman, T., Kokocinski, F., Döhner, H., Bentz, M., Schwäenen, C., and Lichter, P. (2002). Automated screening for genomic imbalances using matrix-based comparative genomic hybridization. *Lab. Invest.*, **82**, 47–60.
- Westfall, P. and Young, S. (2003). *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*. New York: John Wiley and Sons, Inc.
- Wiltshire, T., Pletcher, M. T., Batalov, S., Barnes, S. W., Tarantino, L. M., and Cooke, M. P. *et al.* (2003). Genome-wide single-nucleotide polymorphism analysis defines haplotype patterns in mouse. *Proc Natl Acad Sci*, **100**, 3380–3385.
- Winzeler, E. A., Lee, B., McCusker, J. H., and Davis, R. W. (1999). Whole genome genetic-typing in yeast using high-density oligonucleotide arrays. *Parasitology*, **118**, S73–S80.
- Wolf, C., Flechtenmacher, C., and Dietz, A. *et al.* (2004). p53-positive tumor-distant squamous epithelia of the head and neck reveal selective loss of chromosome 17. *Laryngoscope*, **114**, 698–704.
- Wolff, E., Girod, S., Liehr, T., Vorderwulbecke, U., Ries, J., Steininger, H., and Gebhart, E. (1998). Oral squamous cell carcinomas are characterized by a rather uniform pattern of genomic imbalances detected by comparative genomic hybridisation. *Oral Oncol*, **34**, 186–190.
- Wong, A. K., Biddle, F. G., and Rattner, J. B. (1990). The chromosomal distribution of the major and minor satellite is not conserved in the genus mus. *Chromosoma*, **99**, 190–195.
- Yu, L. C., Moore, D. H. n., Magrane, G., Cronin, J., Pinkel, D., Lebo, R. V., and Gray, J. W. (1997). Objective aneuploidy detection for fetal and neonatal screening using comparative genomic hybridization (cgh). *Cytometry*, **28**, 191–197.
- Yuan, B., Shum-Siu, A., Lentsch, E. M., Hu, L. H., and Hendler, F. J. (1996). Frequent dna polymorphisms exist in inbred cba/j and c3h/hen mice. *Genomics*, **38**, 58–71.
- Yuan, Z. Q., Sun, M., Feldman, R. I., Wang, G., Ma, X., Chen, J., Coppola, D., Nicosia, S. V., and Cheng, J. Q. (2000). Frequent activation of akt2 and induction of apoptosis by inhibition of phosphoinositide-3-oh kinase/akt pathway in human ovarian cancer. *Oncogene*, **19**, 2324–2330.
- Zardo, G., Tiirikainen, M. I., Hong, C., Misra, A., Feuerstein, B. G., Volik, S., Collins, C. C., Lamborn, K. R., Bollen, A., Pinkel, D., Albertson, D. G., and Costello, J. F. (2002). Integrated genomic and epigenomic analyses pinpoint biallelic gene inactivation in tumors. *Nat. Genet.*, **32**, 453–458.
- Zeidler, M., Perrimon, N., and Strutt, D. (1999). The four-jointed gene is required in the drosophila eye for ommatidial polarity specification. *Curr Biol*, **9**, 1363–1372.
- Zeng, H., Chen, Z.-S., Belinsky, M. G., Rea, P. A., and Kruh, G. D. (2001). Transport of methotrexate (mtx) and folates by multidrug resistance protein (mrp) 3 and mrp1: effect of polyglutamylation on mtx transport. *Cancer Res.*, **61**, 7225–7232.

- Zhang, L., Yu, J., Willson, J. K., Markowitz, S. D., Kinzler, K. W., and Vogelstein, B. (2001a). Short mononucleotide repeat sequence variability in mismatch repair-deficient cancers. *Cancer Res.*, **61**, 3801–3805.
- Zhang, W., Remenyik, E., Zelterman, D., E., B. D., and Wikonkal, N. M. (2001b). Genetic progression and clonal relationship of recurrent premalignant head and neck lesions. *Proc Natl Acad Sci U S A*, **98**, 13948–13953.
- Zhou, B.-B. and Bartek, J. (2004). Toward maintaining the genome: Dna damage and replication checkpoints. *Nat Rev Cancer*, **4**, 216 – 225.
- Zhu, C., Mills, K. D., Ferguson, D. O., Lee, C., Manis, J., Fleming, J., Gao, Y., Morton, C. C., and Alt, F. W. (2002). Unrepaired dna breaks in p53-deficient cells lead to oncogenic gene amplification subsequent to translocations. *Cell*, **109**, 811–821.
- Zimonjic, D., Brooks, M. W., Popescu, N., Weinberg, R. A., and Hahn, W. C. (2001). Derivation of human tumor cells in vitro without widespread genomic instability. *Cancer Res.*, **61**, 8838–8844.
- Zweemer, R. P., van Diest, P. J., Verheijen, R. H. M., Ryan, A., Gille, J. J. P., Sijmons, R. H., Jacobs, I. J., Menko, F. H., and Kenemans, P. (2000). Comparative genomic hybridization of microdissected familial ovarian carcinoma: two deleted regions on chromosome 15q not previously identified in sporadic ovarian carcinoma. *Gynecol. Oncol.*, **76**, 45–50.
- Zweemer, R. P., Ryan, A., Snijders, A. M., Hermsen, M. A., Meijer, G. A., Beller, U., Menko, F. H., Jacobs, I. J., Baak, J. P., Verheijen, R. H., Kenemans, P., and van Diest, P. J. (2001). Comparative genomic hybridization of microdissected familial ovarian carcinoma: two deleted regions on chromosome 15q not previously identified in sporadic ovarian carcinoma. *Lab. Invest.*, **81**, 1363–1370.

# Acknowledgements

This thesis is the result of five years of work at the Cancer Research Institute of the University of California in San Francisco. I would like to express my sincere gratitude to all the people who contributed to this work.

In particular, I would like to thank Donna Albertson, who has been my mentor for the past five years. Donna, thank you for all your guidance, support and encouragements throughout this time. Your endless enthusiasm and creativity have always inspired me. No matter how busy you were, you always had time to discuss the latest results, new ideas and so much more. I feel truly honored and privileged to know you and to be able to work under your guidance. Thank you Donna!

I would also like to thank Paul van Diest and Gerrit Meijer. Paul, your support, enthusiasm and positive attitude will always inspire me. Thank you for all your help. Gerrit, thank you for your words of interest and support throughout this time and for your collaboration with us.

Next, I would like to thank Daniel Pinkel for being my mentor. Dan, thank you for sharing your knowledge. I admire your ability to work at the cutting edge of science by being able to combine so many disciplines. I thank you for always taking the time to explain things and to "clear my head" over the years.

Special gratitude goes to Rick Segraves and Bing Huey who have taught me the ins and outs of the lab as well as American life. Thanks! I would also like to express my gratitude to Jane Fridlyand for providing excellent statistical support over the years. I would like to thank Ajay Jain and Taku Tokuyasu for their contributions and continued support. Special thanks also go to Mario Hermsen for introducing me to the world of cancer research, continued collaboration and a good friendship.

Many collaborators have contributed to the work described in this thesis. I would like to thank all of them. Much of this work would not have been possible without their contributions. In particular I would like to thank Norma Nowak and Jeffrey Conroy for providing the human and mouse clone sets. I would also like to thank Brian Schmidt with whom we started collaborating on oral squamous cell carcinoma. Brian, I truly admire your ability to combine an overfull schedule of surgery and patient care with research. Thank you for sharing your knowledge and support. I am looking forward to our continued collaboration in this exciting field.

I am indebted to all of my colleagues, many of whom have since moved on, for their continued support over the years and for creating a wonderful and pleasant atmosphere in the lab. I will name a few and please accept my apologies in advance if I have inadvertently omitted or misspelled your names: Bauke, Bing, Catherine, Dorus, Greg, Hetal, Hua, Joe, Joel, John, Josh, Kate, Ken, Kristin B., Kristin L., Kubilay, Leonore, Nils, Pavla, Pearl, Rick, Serena, Stephanie, Susan, Terri, Tineke, Vanessa, Vita and XiaoXiao. Thank you all for this wonderful experience.

I also would like to thank Frank McCormick and everybody at the UCSF Cancer Center/Cancer Research Institute who has contributed in one way or another to the work presented in this thesis. In particular I would like to thank David Ginzinger (Genome Analysis Core), Karen Kimura (Informatics Core), Randy Davis (Microarray Core), Gregg Magrane (Immunohistochemistry and Molecular Pathology Core) and their excellent staff for rapidly providing their services.

I would like to thank my brother for his tremendous help and interest regarding the layout, design and printing of this thesis. Bedankt, Emile! I am very grateful to my friends and family, in particular, Marianne, Gerard, Anne, Emile, Renske, Opa, Oma, Kelvin, Kin and Melody for their encouragements and support throughout life. Last but not least, I would like to thank Cindy for her everlasting love, her capacity to put me back together when I break down and for all the 'little' things in life.

# List of Publications

## Peer Reviewed

Snijders AM\*, Schmidt BL\*, Fridlyand J, Dekker N, Pinkel D, Jordan RCK, Albertson DG. (2004). Rare amplicons implicate frequent misspecification of cell fate in oral squamous cell carcinoma. *Submitted*.

(\*These authors contributed equally.)

Schmidt BL\*, Snijders AM\*, Fridlyand J, Dekker N, Jordan RCK, Silverman Jr. S, Albertson DG. (2004). Demonstration of oral squamous cell carcinoma second primary clonality using genome wide array Comparative Genomic Hybridization. *Submitted*.

(\*These authors contributed equally.)

Zhang XX, Snijders AM, Zhang X, Seagraves R, Niebuhr A, Albertson D, Yang H, Gray J, Niebuhr E, Bolund L, Pinkel D. (2004). High resolution mapping of genotype-phenotype relationships in Cri du Chat syndrome using array CGH. *Submitted*.

Hermesen M, Snijders A, Guervós MA, Tänzer S, Körner U, Baak J, Pinkel D, Albertson D, van Diest P, Meijer G, Schröck E. (2004). Centromeric chromosomal translocations show tissue-specific differences between squamous cell carcinomas and adenocarcinomas. *Oncogene, in press*.

Snijders AM, Nowak NJ, Huey B, Fridlyand J, Law S, Conroy J, Tokuyasu T, Demir K, Chiu R, Mao J-H, Jain AN, Jones SJM, Balmain A, Pinkel D and Albertson DG. (2004). Mapping segmental and sequence variations among laboratory mice using BAC array CGH. *Genome Research, preliminary accepted*.

Fridlyand J, Snijders AM, Pinkel D, Albertson DG and Jain AN. (2004). Hidden Markov models approach to the analysis of array CGH data. *Journal of Multivariate Analysis.*, **90**, 132-153.

Ishkanian AS, Malloff CA, Watson SK, deLeeuw RJ, Chi B, Coe BP, Snijders A, Albertson DG, Pinkel D, Marra MA, Ling V, MacAulay C, Lam WL. (2004). A tiling resolution DNA microarray with complete coverage of the human genome. *Nature Genet.*, **33**, 299-303.

Snijders AM, Fridlyand J, Mans DA, Seagraves R, Jain AN, Pinkel D and Albertson DG. (2003). Shaping of tumor and drug resistant genomes by instability and selection. *Oncogene*,

22, 4370-4379.

Snijders AM, Nowee ME, Fridlyand J, Piek JMJ, Dorsman JC, Jain AN, Pinkel D, van Diest PJ, Verheijen RHM and Albertson DG. (2003). Genome-wide array based Comparative Genomic Hybridization reveals genetic homogeneity and frequent copy number increases encompassing *CCNE1* in Fallopian tube carcinoma. *Oncogene*, **22**, 4281-4286.

Weiss MM, Kuipers EJ, Postma C, Snijders AM, Stolte M, Vieth M, Pinkel D, Meuwissen SG, Albertson D, Meijer GA. (2003). Genome wide array Comparative Genomic Hybridisation analysis of premalignant lesions of the stomach. *Mol Pathol.*, **56**:293-298.

Weiss MM, Kuipers EJ, Postma C, Snijders AM, Siccama I, Pinkel D, Westerga J, Meuwissen SG, Albertson DG, Meijer GA. (2003). Genomic profiling of gastric cancer predicts lymph node status and survival. *Oncogene*, **22**, 1872-1879.

Weiss MM, Snijders AM, Kuipers EJ, Ylstra B, Pinkel D, Meuwissen SG, van Diest PJ, Albertson DG, Meijer GA. (2003). Determination of amplicon boundaries at 20q13.2 in tissue samples of human gastric adenocarcinomas by high-resolution microarray Comparative Genomic Hybridization. *J Pathol.*, **200**, 320-326.

Jain AN, Tokuyasu TA, Snijders AM, Segraves R, Albertson DG, Pinkel D. (2002). Fully automated quantification of microarray image data. *Genome Res.*, **12**, 325-332.

Snijders AM, Nowak N, Segraves R, Blackwood S, Brown N, Conroy J, Hamilton G, Hindle AK, Huey B, Kimura K, Law S, Myambo K, Palmer J, Ylstra B, Yue JP, Gray JW, Jain AN, Pinkel D and Albertson DG. (2001). Assembly of Microarrays for Genome-Wide Measurement of DNA Copy Number by CGH. *Nature Genet.*, **29**, 263-264.

Cheung VG, Nowak N, Jang W, Kirsch, IR, Zhao S, Chen X-N, Furey TS, Kim U-J, Kuo W-L, Olivier M, Conroy J, Kasprzyk A, Massa H, Yonescu R, Sait S, Thoreen C, Snijders A, Lemyre E, Bailey JA, Bruzel A, Burrill WD, Clegg SM, Collins S, Dhami P, Friedman C, Han CS, Herrick S, Lee J, Ligon AH, Lowry S, Morley M, Narasimhan S, Osoegawa K, Peng Z, Plajzer-Frick I, Quade BJ, Scott D, Sirotkin K, Thorpe AA, Gray JW, Hudson J, Pinkel D, Ried T, Rowen L, Shen-Ong G L, Strausberg RL, Birney E, Callen DF, Cheng J-F, Cox DR, Doggett NA, Carter NP, Eichler EE, Haussler D, Korenberg JR, Morton CC, Albertson D, Schuler G, de Jong PJ, and Trask, BJ. (2001). Integration of cytogenetic landmarks into the draft sequence of the human genome. *Nature*, **409**, 953-958.

Ryan A, Zweemer R, Snijders AM, Hermsen MAJA, Meijer GA, Beler U, Menko FH, Jacobs IJ, Verheijen R, Kenemans P and van Diest PJ. (2001). Comparative Genomic Hybridisation of microdissected familial ovarian carcinoma: two deleted regions on chromosomes 15q not previously identified in sporadic ovarian carcinoma. *Lab. Invest.*, **81**, 1363-1370.

### **Invited**

Snijders A, Segraves R, Pinkel D, Albertson D, (2004). Genomic DNA microarray for Comparative Genomic Hybridization. *Cell Biology: A Laboratory Handbook*. Academic Press,



3rd Edition, *in press*.

Snijders AM, Seagraves R, Blackwood S, Pinkel D, Albertson DG. (2004). BAC Microarray-based Comparative Genomic Hybridization. *Methods in Molecular Biology*, **256**, 39-56, Humana Press.

Snijders AM, Pinkel D, Albertson DG. (2003). Current status and future prospects of array based Comparative Genomic Hybridization. *Briefings in Functional Genomics and Proteomics*, **2**, 37-45.

Weiss MM, Hermsen MAJA, Snijders A, Buerger H, Boecker W, Kuipers EJ, van Diest PJ, Meijer GA. (2003). Comparative Genomic Hybridization in pathology. *Molecular Biology in Cellular Pathology*. John Wiley and Sons, Ltd.

Snijders A, Pinkel D, Albertson D. (2002). Preparation of BAC DNA for CGH by ligation-mediated PCR. *DNA Microarrays; A Molecular Cloning Manual*. CSHL press.

Snijders AM, Meijer GA, Brakenhoff RH, van den Brule AJC, van Diest PJ. (2000). Microarray techniques in pathology: tool or toy? *J Clin Pathol-Mol Pa* **53**, 289-294.



# Curriculum Vitae

The author was born on June 21, 1977 in Nunhem, The Netherlands. After finishing high school (VWO) in 1995 at Scholengemeenschap St. Ursula, Horn, he studied Medical Biology at the Free University (Amsterdam, The Netherlands). His major required that he conducted two internships at different departments. As a consequence, he conducted his first internship at the Department of Pathology at the Free University Medical Center (Amsterdam, The Netherlands) where he used fluorescent *in situ* hybridization and immunohistochemical techniques to gain insights in human colorectal tumors (mentor: Dr. M.A.J.A. Hermsen). Then he worked for several months at the Departments of Pathology and Gynaecology at the Free University Medical Center (Amsterdam, The Netherlands) where he initiated a study to investigate DNA copy number aberrations in familial ovarian carcinoma (mentor: Prof. Dr. P.J. van Diest). Then he conducted his second internship at the Cancer Research Institute of the University of California San Francisco (San Francisco, USA) where he developed a technology for the preparation of spotting solutions for array Comparative Genomic Hybridization (mentor: Prof. Dr. D.G. Albertson).

He obtained his Master's of Science degree in Medical Biology in 1999. In 2000, he began his PhD study under guidance of Prof. Dr. D.G. Albertson, Prof. Dr. P.J. van Diest, Prof. Dr. G.A. Meijer and Prof. Dr. D. Pinkel. The work described in this thesis was performed at the Cancer Research Institute of the University of California San Francisco (San Francisco, USA). In 2004 he received the President's award for excellence for young investigators at the XXII International Congress of the International Society for Analytical Cytology (ISAC). His graduation will take place at the University of Utrecht (Utrecht, The Netherlands). From 2005, the author will continue to work with Prof. Dr. D.G. Albertson at the Cancer Research Institute of the University of California San Francisco.