Genomic Analysis of Mouse Tumorigenesis

by

Mandy Chi-Mun Tam

B.Sc. Biochemistry
University of Toronto, 2000

Submitted to the Department of Biology in Partial Fulfillment
of the requirements for the degree of

Doctor of Philosophy

at the

Massachusetts Institute of Technology

September 2006

Signature of author: _____
Department of Biology
August 15, 2006

Certified by: _____
Tyler Jacks
Professor of Biology
Thesis Supervisor

Accepted by: _____
Stephen Bell
Professor of Biology
Chairman, Biology Graduate Committee

Genomic Analysis of Mouse Tumorigenesis

by

Mandy Chi-Mun Tam

Submitted to the Department of Biology on August 21, 2006 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Biology

ABSTRACT

The availability of the human and mouse genome sequences has spurred a growing interest in analyzing mouse models of human cancer using genomic techniques. Comparative genomic studies on mouse and human tumors can be valuable in two major ways: in validating mouse models and in identifying genes that are common to mouse and human tumorigenesis. Many analytic tools have emerged in recent years for human genome mining. Some of these tools have been translated to the murine versions. The work in this thesis described the application of two new whole-genome analytic techniques to study mouse tumorigensis: Representational Oligonucleotide Microarray Analysis (ROMA) for tumor DNA copy number asessment and single nucleotide polymorphism (SNP) genotyping using the SNaPshot™ system (Applied Biosystems) to detect loss of heterozygosity (LOH) in mouse tumors. The murine version of ROMA was tested on DNA from early-stage $Kras^{G12D}$-derived lung cancers and metastatic retinoblastoma in mice with retinal-specific *Rb* and *p130* deletions. We were interested in identifying the additional genetic lesions that got positively selected during tumorigenesis of these mice. Several recurrent chromosomal copy number gains and losses were observed in the DNA of $Kras^{G12D}$-derived lung tumors. In addition, a focal amplification of the murine *N-Myc* locus was detected in the metastatic retinoblastomas, demonstrating the capability of ROMA to detect copy number changes at a single-gene resolution. For genome-wide allelotyping, a panel of 147 mouse SNPs were individually validated in 129S4/SvJae vs. C57BL/6J strains and were chosen as markers in the genotyping panel. We worked out a multiplex protocol to genotype the SNPs in an efficient manner. Through this protocol, we generated low-density global LOH maps of lung tumors from mice expressing $Kras^{G12D}$. LOH that spanned entire chromosomes was seen in a subset of the tumors. A loss of the wild-type p53 allele was also observed in some cases.

Thesis Supervisor: Tyler Jacks, Ph.D.

Position: Professor of Biology

ACKNOWLEDGEMENTS

The biggest reward in graduate school was the opportunity to meet many driven, bright, and interesting people. It is from them I learned much about science and about life. Biology is a study of life after all.

I'd like to thank my advisor, Tyler Jacks, who guided my journey with much patience and understanding. It was a privilege to be able to learn from someone who leads with scientific insights, wisdom, and compassion.

I am thankful to my thesis committee members and co-workers for their scientific advices. I was also fortunate to have collaborators who made this thesis work possible. I especially want to thank Alejandro Sweet-Cordero, David MacPherosn, David Mu, David Ginzinger, and all the technicians at the UCSF cancer center genome core for their assistance and guidance.

I am grateful to have a group of classmates, roommates and friends who pushed my limit, kept me grounded, and made me laugh. I truly believe every person I met in this journey has made an impact on me. In particular, I'd like to thank my two classmates and good friends Ishara Mills and Mark Rosenzweig for their support and help in putting this thesis together and throughout graduate school.

Finally, I want to acknowledge my parents, past teachers, and college friends, who shaped or shared my values, and supported my decision to pursue graduate study.

Table of Contents

Chapter 1

Introduction:

Analysis of Genetic Alterations in Cancer

Viewing cancer from an evolutionary perspective

Beneath the diversity of life is the dynamic process of change that alters chances for living. Some changes are ephemeral while others, such as those in the form of genetic mutation, can last for generations. Over time, only heritable traits can endure the test of natural selection, in which mutation is the key engine. The result is variations among life forms: the different species and the different individuals within a species. Some will thrive and some will not.

The disease cancer has diverse manifestations: in its onset, pathology, malignancy and therapeutic response. This variation can probably be explained most fittingly from a Darwinian perspective. Played out by evolutionary rules, mutation and selection operate in a micro-scale to govern cancer development in the body. Each cancer is a clone of mutated cells selected for the ability to multiply and grow in an environment that is short of nutrients, oxygen and other survival factors. As each clone proliferates, successive mutations continue to create subclones with enhanced growth advantages. Neoplastic cells progress through malignant stages in these waves of clonal expansion. It is the inherent randomness of mutation that gives cancer its heterogeneous and often unpredictable expressions. Given the intrinsic mutability of the genome, cancer is perhaps an inevitable baggage of species evolution. Retention and refinement of genes that regulate cell growth, cell fate, cell migration, and cell-cell interaction allow our species to exist and flourish (Hanahan and Weinberg 2000). The same genetic networks, however, when deregulated by the inexorable force of mutation, have the potential to

destroy an individual in the form of cancer. Cancer is essentially an uninvited hitchhiker in the evolutionary probability game.

Cancer research aims to decode the micro-evolutionary game rules and to develop strategies to play against the odds. The analysis of mutations in tumors is fundamental to our understanding of the elements of the disease. The past few decades mark important developments in our understanding of molecular biology and technical ability to perform genetic analyses. Numerous oncogenes and tumor suppressor genes have been identified. Some become critical targets for therapy. With the recent emergence of genomics, cancer genetics has moved from single gene analyses to whole-genome diagnoses of the multiple mutations that exist. This development has allowed researchers to more fully capture the variable genetic landscape that is characteristic of cancer. This thesis describes efforts to identify genetic alterations in mouse models of human cancer by implementing some of the latest techniques in genome analysis. The analysis focuses on lung cancer, the current leading cause of cancer deaths worldwide, and touches on retinoblastoma, a childhood eye cancer. This first chapter will introduce readers of this thesis to various cancer genome analytic tools. Chapters 2 and 3 will summarize results generated by two complementary genomic techniques in the study of mouse tumorigenesis. Finally, the implications of this research will be discussed in the concluding chapter.

Mutations as underlying cause of cancer

While the concept that cancer is caused by underlying mutations is well known today, it is interesting to look back at how this notion arrived. Mankind recognized and named the

disease cancer over two thousand years ago. Prior to the emergence of experimental medicine, the cause of cancer was widely believed to be an imbalance of bodily humors. The premise was replaced in the 18th century by the lymph theory, which assumed cancer grew out of abnormal lymph fluid. In 1890, David von Hansemann made the first report of abnormal mitoses in tumors, suggesting a genetic cause of cancer (Shimkin 1977). In 1914, Theodor Boveri postulated the somatic mutation theory, which identified chromosomal abnormalities as possible culprits that caused cells to adopt cancerous properties (Manchester 1995). The theory, however, remained a conjecture for a few more decades due to limits in cellular and molecular genetics techniques at the time. Other prevailing ideas from the same period proposed cancer originated from trauma, viruses, or environmental factors. The trauma theory got disproved in the late 1920's; subsequently, researchers showed that cancer viruses disrupted expression of genes, while radiation and numerous chemical carcinogens were found to act by inducing mutations in the genome (Shimkin 1977). In 1960, a major step in verifying Boveri's mutation hypothesis came when Nowell and Hungerfold discovered in patients with chronic myeloid leukemia (CML) the Philadelphia chromosome, the first recurrent chromosomal abnormality associated with a cancer (Nowell and Hungerford 1960). In other cancers, while the probabilistic nature of mutation has often made it difficult to isolate specific genetic aberrations, retrospective surveys of data have eventually revealed a non-random pattern of genetic changes in cancers of most organs. The notion that cancer evolves through a selection of particular mutated genetic elements has gotten accepted beyond doubt. Some of these genetic constituents were later experimentally identified and broadly classified as oncogenes, for genes that enhance proliferation,

growth, and differentiation when activated, and tumor suppressor genes, for genes that lose their normal regulatory roles when inactivated in cancer.

Early cancer gene discovery tools

Methods to identify disease genes fall into three categories: cytogenetic mapping, physical mapping, and linkage mapping. In the early days of cancer research, cytogenetics was the most feasible way to study genomic alteration in cells under a microscope. Chromosome banding techniques have enabled the discovery of numerous cancer-related chromosomal aberrations (Mitelman 2000). Recurrent translocations have been particularly useful in uncovering numerous oncogenes such as Abl in BCR-Abl and c-Myc in IgG-cMyc fusions of CML and Burkitt's lymphoma. Functional genomic screens such as the in-vitro transformation assay also helped to identify genomic DNA fragments containing oncogenes (Shih et al. 1981). Early genetic maps with restriction fragment length polymorphism (RFLP) and simple sequence length polymorphism (SSLP) markers have allowed the discovery of tumor suppressor genes through linkage mapping in familiar cancer syndromes and loss of heterozygosity (LOH) mapping in tumors of individuals.

Reflections from past studies

A recent review has summarized the many cancer gene mutations known today (Futreal et al. 2004). In that summary, it is clear that cancer genetic alterations appear in virtually all types but chromosomal aberrations are the predominance. Among the recurrent

chromosomal changes, the target genes involved still remain to be identified. The Mitelman database is a catalog of cancer chromosomal aberrations (Mitelman et al. 2006). As of March, 2006, chromosomal aberrations have been documented in over 49,000 cancer samples in the forms of monosomy, trisomy, balanced and unbalanced translocations, and amplifications and deletions in various scales. In the numerous reports without observed chromosomal alterations, mutations at the nucleotide level such as microsatellite and other repeat instabilities and point mutations are often seen. While carcinomas constitute over 80% of all human cancers, carcinoma-related genetic alterations only represent 20-30% of the cataloged cases in both databases (Futreal et al. 2004; Mitelman et al. 2006).

It is appearing that the many classes of mutations in cancer may require complementary approaches for discovery, and that many more cancer genes, especially those involved in the carcinoma development, remain to be discovered. With the availability of the human genome sequence, various new analytical tools have been developed to characterize the cancer-related genetic changes at a genomic level. The tremendous amount of information within a genome is best mined with tools with high throughput and resolution, the two grounds on which new technologies are competing.

Evolving technologies to analyze the cancer genome

Comparative Genome Hybridization (CGH)

CCH involves simultaneous hybridization of differentially labeled test and reference DNA to obtain relative copy number information along a chromosomal position coordinate. Copy number alterations can reflect aneuploidy, unbalanced translocations, gene amplifications or deletions. Prior to genome sequencing, chromosomal positions were cytogenetically traced to bands on normal metaphase spreads. Genome sequencing has allowed the use of arrayed DNA stretches with known positions as mapping coordinates, thus raising the cap of resolving power (Albertson and Pinkel 2003). This section will summarize the basics and utilities of CGH in cancer genomics research.

CGH platforms

Arrays of large-insert clones

Initial array CGH platforms utilized large-insert genomics clones as the probing elements. These include BAC (Bacterial Artificial Chromosomes), YAC (Yeast Artificial Chromosomes), PAC (P1 Artificial Chromosomes), and cosmids (Solinas-Toldo et al. 1997; Pinkel et al. 1998). The initial study by Pinkel et al. (1998) showed the ability of BAC microarrays to detect difference in X-chromosome number in sex mistached samples and known copy number increases on chromosome 20 in breast cancer cell lines. More importantly, they were able to detect a novel deletion on 20q chromosome arm, showcasing the potential of BAC arrays to find changes that were missed by older methods.

Array spotted with ~2400 BAC and P1 clones with an average ~1Mb resolution for the human genome was reported in (Snijders et al. 2001). To achieve better resolution, a tiling array with 32433 overlapping BAC clones was constructed (Ishkanian et al. 2004). A sub-Mb resolution was reported, allowing detection of a 300kb validated amplicon on 13q12.2 in a colorectal cell line and a 240kb validated deletion in a breast cancer cell line. While the power of such tiling BAC array is clear, the manufacturing of such array requires large-scale management of clones, making the technology hard for most labs to adopt. PCR strategies including degenerate oligonucleotide PCR (DOP-PCR) and ligation-mediated PCR have been successfully employed to amplify BACs before spotting (Hodgson et al. 2001).

Arrays of cDNA

Microarrays originally used for genome-wide gene expression analysis have been used for CGH (Pollack et al. 1999). Because genomic DNA is more complex than RNA representations, CGH requires a platform with a higher minimum sensitivity than one for gene expression analysis. The feasibility of cDNA and EST-based CGH was proved by Pollack et al. in 1999. Using Turner syndrome patient and sex-mismatched samples, they were able to detect gene copy number differences on X and Y chromosomes. In addition, they confirmed the known *ERBB2* and *MYC* amplifications and *p53* deletion in established cell lines. At a genomic level, they observed 62% highly amplified genes also have elevated expression, backing the case for a gene dosage effect on gene expression.

However, the signal:noise ratio was poor, making averaging of signals over several probes necessary, which essentially limited resolution.

Arrays of oligonucleotides

Oligonucleotides can also be used as arraying elements. This is made possible by methods that allow precise in-situ synthesis of oligonucleotides include photolithography (Affymetrix, Nimblegen) and inkjet printing (Agilent). Such printing technologies have also made the manufacturing of oligonucleotide arrays scaleable.

Theoretically, oligonucleotide array probes can be designed for any sequence, enabling high-density coverage of any region of interest. In practice, resolution is a function of signal to noise. Because of their short lengths, hybridization signal on each oligonucleotide is generally lower than on a BAC probe. As in cDNA arrays, averaging of signals over a few adjacent probes is necessary to obtain reliable calls, which in effect decreases resolution.

One of the first successful implementations of oligonucleotide arrays was reported by (Barrett et al. 2004). The investigators successfully performed genome-wide CGH on Agilent Technology's ~17k oligonucleotide array originally designed for gene expression analysis. In addition, using probes designed specifically for CGH on chr x, 18, and 17 they were able to detect single copy change as verified in known chromosome x number difference in various cell lines. Later, a whole-genome array made by Agilent was reported to achieve a ~70kb average resolution (Brennan et al. 2004). Affymetrix also has

developed oligonucleotide arrays for CGH. Using ASO arrays designed for SNP detection, LOH and copy number changes could be used detected simultaneously (Bignell et al. 2004; Zhao et al. 2004). While most other platforms involve co-hybridization of tumor vs. normal DNA on each array, the Affymetrix chip is a single-channel array, requiring hybridization of a set of normal samples separately as reference. Non-commercially made oligonucleotide CGH arrays have also been reported(Carvalho et al. 2004; van den Ijssel et al. 2005). Recently, two oligonucleotide arrays were made and tested by Nimblegen: 1) a genome-wide array with 6kb median spacing and 2) a tiling oligonucleotide array covering selected regions at a sub-kb interval (Selzer et al. 2005), further demonstrating the power of oligonucleotide CGH arrays.

Arrays with other elements

ssDNA has been reported as probes in exon array CGH (Dhami et al. 2005). As a proof of principle, Dhami et al. tested an array with ssDNA probes to 162 exons for 5 human genes. Compared to arrays made with other probing elements, the ssDNA array had an enhanced signal: noise ratio and thus have a sensitivity of ~2x higher. It remains to be seen whether this method can be scaled to cover the whole genome.

Applications in cancer research

Microarray formats provide the high-throughput capabilities for genome-wide studies. Some of the pioneering cancer studies are discussed above. Genome-wide CGH analysis has become increasingly routine in cancer research. A recent study by Tonon et al. on

lung cancer showcased the power of array CGH in detecting disease loci. Using oligonucleotide and cDNA arrays made by Agilent to characterize human lung cancer samples and cell lines, numerous recurrent focal copy number changes could be seen, some were below 0.5Mb (Tonon et al. 2005). The authors' comparison of the CGH data with a gene expression analysis enabled them to further pinpoint critical candidates, which included p63 in squamous cell carcinoma. Beyond the use of CGH data in candidate gene discovery, it is worth mentioning its application in tumor classification, as has been done to distinguish renal cancer (Wilhelm et al. 2002) and multiple myeloid subtypes (Carrasco et al. 2006).

The case for genome complexity reduction

The human genome is very complex: over 50% of the genome sequence is comprised of repeats whereas only ~5% is protein coding (Lander et al. 2001). The abundance of these repeats, plus low-level shared sequences such as those by gene family members, pose a challenge for specific hybridization. Many hybridization-based genetic analysis protocols employ a genome complexity reduction step. In the classic case of Southern blotting, genomic DNA is fragmented and separated by electrophoresis prior to hybridization. For genomic analysis, PCR based methods are used. One is to degenerate oligonucleotides to random prime and amplify the genome (Telenius et al. 1992; Kuukasjarvi et al. 1997). The alternative method is restriction-enzyme based, which is more reproducible (Lucito et al. 1998). In the latter case, the resulting complexity of the representation can be controlled by the choice of restriction enzyme; a less frequent cutter would generate a

less complex representation. Theoretically, a low-complexity representation (LCR) would improve hybridization by enhancing signal: noise ratio and reducing hybridization times.

A protocol of preparing LCR was described in (Baldocchi and Flaherty 1997). The method involves digestion of test and reference genomic DNA with *Bgl*II, a 6bp cutter, and linker-based PCR-amplification. As PCR selectively amplifies smaller restriction fragments, only a ~2.5% low-complexity representation of the genome is made by this process.

The usefulness of LCRs in genome-wide copy number analysis was first demonstrated in subtractive hybridization, in which a complexity reduction step is essential for mammalian genome analyses; the method of LCR subtraction is called Representational Difference Analysis (RDA) (Lisitsyn and Wigler 1995). RDA of tumor and normal genomes have been successfully performed to identify copy number alterations, including PTEN loss in multiple cancers (Lisitsyn et al. 1995; Li et al. 1997; Hamaguchi et al. 2002; Mu et al. 2003). Based on these successes, the next conceivable step was to test the use of LCR in array-based hybridization.

ROMA –an alternative way to perform genome-wide CGH analysis

Representational Oligonucleotide Microarray (ROMA) involves hybridization of LCRs on an array for copy number comparison. In a pilot experiment, 1-2K oligonucleotide arrays for used to probe for BglII representations (Lucito et al. 2000). They were able to demonstrate the preservation of original gene ratios in the LCRs, and that results were

reproducible and had good signal to noise ratio. In addition, in the making of LCRs, small amount of samples could be amplified and that polymorphism at restriction site could be reflected in parallel. In a later experiment, a denser microarray (Nimblegen) with 80K oligonucleotide probes was tested (Lucito et al. 2003). An average genome-wide resolution of 30kb was achieved, and in some area, the resolution was as high as 15kb (Lucito et al. 2003). ROMA thus provides a high-resolution alternative to perform CGH studies (Jobanputra et al. 2005).

Application considerations

In choosing an array platform for CGH study, several factors need to be considered: the quantity and quality of DNA available, the platform performance and the cost and accessibility of the platform. In CGH, there is no need for culturing cells to prepare karyotypes as required for cytogenetic analysis, suggesting even DNA from archival tissues can be analyzed. Quantity and quality of DNA will determine what platforms can be used. BAC arrays can be performed with ~300ng of DNA while cDNA and oligonucleotide arrays require a few micrograms of starting materials, although whole-genome amplification can be performed at stake of an increased experimental cost (Lage et al. 2003). In addition, suboptimal quality of DNA such as those obtained from archival tissues may be incompatible with some protocols such as ROMA. It is often debated which array CGH platform will dominate the future. Resolution is likely the most important determining factor. Other factors such as manufacturability will also affect dominance. On those two grounds, Ylstra et al. recently suggested the future will likely belong to oligonucleotide arrays (Ylstra et al. 2006). Whether their prediction is correct

or not, the increasingly powerful array CGH technology overall will likely yield many exciting data in the cancer genomics field.

Single nucleotide polymorphism (SNP) genotyping

While CGH enables high-resolution physical mapping of copy number alterations in cancer, allelic mapping with genetic markers can also be informative. Genome sequencing has led to the discovery of a vast number of SNPs. To date, over 12 million SNPs in the human genome are referenced in the National Center for Biotechnology Information (NCBI) dsSNP database build 126 (http://www.ncbi.nlm.nih.gov/projects/SNP/). The dense coverage has made SNPs the molecular marker of choice in genetic mapping. The following section will discuss the use of SNP genotyping in cancer research, in genome-wide and regional candidate mapping studies. A description in the different genotyping methods will then follow.

Applications in cancer research

Genome-wide loss of heterozygosity (LOH) mapping

LOH maps can be used to infer tumor suppressor gene locations. In humans, most LOH events are not to be associated with copy number changes and would be missed by CGH (Huang et al. 2004; Beroukhim et al. 2006). LOH can be identified by analyzing SNPs within the tumor and normal DNA of the same individual. Early efforts to use SNPs for high-density LOH mapping included studies on lung cancer, neurofibromatosis type 2, and esophageal adenocarcinomas (Lindblad-Toh et al. 2000a; Mei et al. 2000). Regions

of allelic imbalance could be identified in each case. Recently, using SNP arrays as a detection platform (described in a later section) and algorithms to integrate signal intensities to allelic calls, copy number information can be viewed along the LOH data, allowing the distinction of copy neutral LOH, copy number loss associated LOH, or copy number gain associated allelic imbalances (Bignell et al. 2004; Zhao et al. 2004; Zhao et al. 2005). The power of bioinformatics has also allowed LOH regions to be inferred from tumor samples without paired normal DNA using high-density SNP genotyping data (Beroukhim et al. 2006); such would be impossible without a high-resolution mapping tool. Aside from identifying candidate tumor suppressor gene region, LOH data can also be used for tumor classification, as shown by the LOH-based clustering of non-small cell lung cancer vs. small cell lung cancer samples (Janne et al. 2004).

Genome-wide mapping of cancer susceptibility loci in population studies

Aside from LOH mapping, SNP genotyping is also a powerful tool in linkage or linkage disequilibrium studies that attempt to map disease genes using a family-based approach. Genome-wide mapping can be enhanced by the availability of high-throughput, high-density SNP genotyping techniques. A traditionally popular marker set for mapping is the ABI Prism Mapping Set (Applied Biosystems), which consists of microsatellite markers at 10cM apart. Genome-wide SNP mapping panel such as the Affymetrix GeneChip (discussed below) allows simultaneous anlaysis of 10K SNPs, which means markers are spaced at 0.34cM on average. A linkage study on prostate cancer has compared the used of SNPs and microsatellite markers in mapping(Schaid et al. 2004). The denser SNP markers indeed allowed the authors to get a better linkage resolution and identify more

linkage peaks on multiple chromosomes. Genome-wide SNP based linkage studies have also been performed in other familial cancer cases including hereditary mixed polyposis syndrome (HMPS) (Cao et al. 2006) and chronic lymphocytic leukemia (CLL) (Sellick et al. 2005). For non-familiar cases, studies done by genome-wide SNP mapping of disease susceptibility loci have been performed on breast cancer (Ellis et al. 2006), Bloom syndrome and hereditary nonpolyposis colorectal cancer (HNPCC) (Mitra et al. 2004). Candidate loci were identified in each case; in the Bloom syndrome study, a single locus TSC0754862 was pinpointed (Mitra el al. 2004).

Narrowing down candidate regions with SNPs

Aside from genome-wide mapping, the high density of SNPs makes them the ideal genetic marker for narrowing down candidate regions. One candidate region that remains to be delineated is chromosome 3p, where LOH in multiple areas have been implicated in various cancers including over 90% of all lung cancers (Zabarovsky et al. 2002). While most older studies have approached the delineation problem with microsatellite markers, attempts to identify candidate tumor suppressor genes on chromosome 3p by SNP-based LOH mapping have begun (Tai et al. 2006). A linkage study that aims to scale down candidate tumor susceptibility region on 3p has also been performed on prostate cancer patients (Larson et al. 2005). In addition to delineating large chromosomal regions, gene specific studies to associate particular SNPs with a functional role in cancer have also proved to be informative. One recent study of such kind was performed on *CHK2* in order to identify SNPs that can affect breast cancer susceptibility (Einarsdottir et al. 2006).

SNP genotyping approaches

The potential of exploiting SNPs as markers has stimulated a multitude of imaginative approaches to genotype them (Syvanen 2001; Engle et al. 2006). The choice of a genotyping protocol depends largely on the need of the research and the resources available. As discussed above, various types SNP-based studies can be performed in cancer research. A genome-wide study will put high throughput as top priority while a regional study will benefit from a highly flexible assay in order to genotype the SNPs of choice. The following section is a discussion on the scientific principles behind current SNP genotyping methods. Accuracy and robustness of an assay depend largely on the underlying reaction biochemistry. Then, the format and readout of the assay will determine what instruments are necessary and thus affect the ease of use, throughput, and cost. My discussion will be therefore divided into these two parts. In practice, because of the immense value of SNP genotyping in biomedical research, many assays have been commercialized. Table 1 is a summary of the working principles behind some commercial assays.

Table 1: Decoding commercial buzzwords in SNP genotyping

Assay tradename	Company	Working Priniciples*
GeneChip	Affymetrix	ASO array hybridization
Genorama	Asper	APEX
GoldenGate	Illumina	OLA+ASO+microbeads on microarray
Invader	Third Wave Technology	Invasive cleavage
iPLEX	Sequenom	SBE+MALDI-TOF
LightCycler	Roche	FRET
MassEXTEND	Sequenom	primer extension+MALDI-TOF
PinPoint	Applied Biosystems	primer extension+MALDI-TOF
SNaPshot	Applied Biosystems	SBE+capillary electrophoresis
SNPlex	Applied Biosystems	OLA+capillary electrophoresis
SNPstream	Orchid/Beckman	SBE+microarray

* ASO=allele-specific oligonucleotides
APEX=array-based primer extension
OLA=oligonucleotide ligation assay
SBE=single-based-extension
FRET=fluorescent resonance energy transfer
MALDI-TOF=matrix assisted laser desorption ionization time of flight

Hybridization

Allele-specific oligonucletide (ASO)

In differentiating SNP alleles, two ASO probes are used, each with a different allele of the SNP that is usually in the central region. Probe binds stably to the match allele but less so to the mismatch (Saiki et al. 1988).

Enzymatic approaches

Restriction enzyme

The sensitivity of restriction endonucleases to distinguish short and defined sequences can be exploited. Restriction site length polymorphism (RFLP) is the historical method for genotyping SNP. At where a SNP changes a restriction enzyme site, a different digestion pattern can be seen(Syvanen 2001).

Allele-specific amplification (ASA) or Primer extension

This is a DNA polymerase-based method using two probes with a discriminating base at or near the 3'ends. When primers match the target, Taq polymerase catalyzed extension can occur (Waterfall and Cobb 2001).

Single-base extension (SBE)

In SBE or mini-sequencing, a primer is designed to anneal immediately upstream to the base of the SNP. A polymerase reaction is performed to extend the primers by one base with didedoxynucleotides at the SNP site (Sokolov 1990; Syvanen et al. 1990).

Combined hybrdization-based/enzymatic approaches

Oligonucleotide-ligation assay (OLA)

This assay involves a pair of probes with an allele-discriminating base at one end (either 5' or 3') and another oligonucleotide ending at the base adjacent to the SNP. Ligase mediates joining of the oligonucleotide to the probe that matches the allele(Landegren et al. 1988).

Invasive cleavage assay

The assay employs the use of Flap endonuclease (FEN), an enzyme that recognizes and cleaves the 'flap' that results from the binding of two overlapping oligonucleotides to the same target DNA with perfect match. To exploit this property of FEN to discriminate SNP, three oligonucleotides are employed: one pair of probes containing an internal allele-discriminating base, and an 'invader' oligonucleotide that can bind to target sequence on the 3' side of the SNP. When there is a perfect match, the 'flap' of the probe will get cleaved (Olivier 2005).

Comments on the different biochemistries

The accuracy of hybridization approaches depends largely on binding specificity. As such, attempts have been made to use probes with special nucleotides that can bind to complementary DNA tighter. The increase in stability can improve the allele-discriminating ability of the probe. TaqMan MGB probes that bind to minor groove of

DNA is one example (Kuimelis et al. 1997). Syntheic nucleotide analogs including peptide nucleic acid (PNA) (Ross et al. 1997) and locked nucleic acid (LNA) (Orum et al. 1999) have also been used. PNAs are analogs with uncharged polyamide backbones (Ross et al. 1997) while LNAs contain an extra 2'-O, 4'-C-methylene bridge on the ribose ring of the nucleotide (Orum et al. 1999).

For the enzymatic or combined approaches, accuracy much depends on the fidelity of the enzyme. DNA polymerases for SBE are very accurate. SBE assays are already used widely with various platforms e.g.(Nikiforov et al. 1994; Shumaker et al. 1996; Syvanen 1999), while the newer approaches --OLA and invasive cleavage -- are gaining acceptances as well.

Detection principles

The basis of detection dictates what assay platforms will be required. The two in conjunction affect how sensitive and quantitative the assay is.

Fluorescence

Fluoresence allows quantification and differentiation of alleles. In SBE, different fluoroscein labeled nucleotides can be used for incorporationg. Alternatively, targets can be fluorsencently labeled and amount of binding is measured (such as in 2-D microarrays). On the flip side, probes fixed on different fluorescently labeled microbeads is an alternative arraying format.

FRET

When two fluorophores with overlapping excitation and emission spectra are in close proximity, FRET can occur. FRET is the principle behind TaqMan (Applied Biosystems) probes (Livak et al. 1995; Livak 1999) and Molecular Beacon probes, and in LightCycler assay (Roche). In a TaqMan assay, two probes with different fluorescent dyes at 5' and a 3' quencher are used. Each probe has a discriminating base for each SNP allele. During PCR, probe with a mismatch will be displaced without cleavage while the matched probe gets cleaved, giving out fluorescence signal that gets monitored in real time(Livak et al. 1995; Livak 1999). In a Molecular Beacon assay, two probes with different fluorescent dyes at 5' and a 3' quencher are also used. Molecular Beacons are hairpin probes that contain a sequence complementary to target DNA (Tyagi and Kramer 1996; Tyagi et al. 1998). When the probe binds to perfectly matched target, the hairpin opens up to give up fluorescence. The LightCycler (Roche) assay is similar to TaqMan, but instead of using a single probe, LightCycler uses two different labeled probes binding to adjacent DNA sequences and one contains an allele-specific base (von Ahsen et al. 1999).

Fluorescence polarization (FP)

A DNA binding dye such as SYBR Green can be included in the reaction to detect formation of the product. Fluoresence polarization (FP) can also be used. The method uses polarized light to excite a fluorophore. The direction of emission depends on mass of the molecule, making it able to monitor the change in product size (Germer et al. 2000).

Mass spectrometry (MS)

Mass changes can be measured by matrix assisted laser desorption/ionization time-of-flight (MALDI-TOF) spectrometry. MALDI-TOF MS is highly sensitive and can resolve the smallest nucleotide difference, which is that of 9Da between A and T (Edwards et al. 2005). Thus, MALDI-TOF MS can assay which ddNTP was added to an SBE primer. In addition, SBE primers of varying lengths can be resolved, enabling multiplexing of reactions (Ross et al. 1998; Bray et al. 2001).

Chemilluminescence

Pyrosequencing is one example of chemilluminescence-based method (Langaee and Ronaghi 2005). It is a way to perform SBE. It is a system that measures the released pyrophosphate during addition of nucleotide, added one by one in specific order. The pyrophosphate is converted to ATP by ATP sulfurylase, and light is generated from ATP by luciferase. Degradation of the added nucleotide by apyrase frees up the template for next nucleotide addition round.

Chemogenic signal

Hapten-labelled nucleotides can be used for SBE. Hapten can be recognized by antibodies conjugated by enzyme that catalyzes light product formation like in an ELISA (Friedhoff et al. 1993).

Nanoparticles

Quantum dots

Quantum dots are microfabricated nanoparticles that can be synthesized to emit very bright fluroscence at any wavelength, as emission is dependent on their sizes (Waggoner 2006). In SNP detection, the proof-of-concept use of quantum dots in place of traditional fluorophore has been demonstrated (Han et al. 2001). Quantum dots were able to increase detection sensitivity and multiplexing capability of assay (Xu et al. 2003).

Gold nanoparticles

The intense scattering of absorbed light by gold nanoparticles due to surface plasmon resonance can also be a way to enhance detection signal(Taton et al. 2000). Similar to quantum dots, gold particles have the potential to replace fluorophores to improve SNP detection sensitivity (Taton et al. 2000).

Assay platforms

Detection principle dictates the choice of platforms. The choice of platform determines the fixed cost and affects sensitivity and accuracy of detection. The accessibility of specific platforms can limit the choice of a researcher.

Solid phase:

Dot blots and reverse dot blots

They are the classic methods for hybridizing ASO (Saiki et al. 1988). In a traditional dot blot, target DNA is bound on nitrocellulose or nylon filters and is probed with radiolabeled allele-specific oligonucleotides. Reverse is done in reverse dot blots. Thermodynamics of binding between nucleotide strands is not only affected by the

single-base difference but also the surrounding sequence. Hybridization conditions differ for different probe-target pairs. The use of tetramethylammonium chloride (TMAC) as the hybridization solution has greatly increased the thermal stability range of oligonucleotides, allowing a universal hybridization condition (current protocols of molecular biology).

Microarrays

Microarrays allow simultaneous analysis of a large number of markers . Classicaly these are 2-D glass slides containing bound DNA probes in the form of fixed arrays. Data are obtained from scanned images. Microarrays can be used as a physical platform for hybridization or as a reaction platform for in-situ biochemistry.

The first usage essentially evolves from the reverse dot blotting principle(Southern et al. 1992). One subtype is to attach ASO probes to the glass slide and allow hybridization of fluorescently labeled DNA with the SNP targets (Ranade et al. 2001). To circumvent the thermodynamic differences between different probe-target pairs, more than one probe per SNP can be used. For instance, the Affymetrix GeneChip system uses tens to hundred of ASO probes for each SNP. The other subtype of hybridization-based microarrays is to use different bound probes to fish for the SNP products resulting from one of the enzymatic reactions described. A new array format using microelectrodes to immobilize DNA has been developed by Nanogen. By controlling the current, binding and washing of DNA can been expedited. Application of such chip in hybridization-based SNP detection has been proven to be robust (Gilles et al. 1999).

Aside from being a hybridization scaffold, microarrays can also be used as a platform to carry out SBE and OLA. SBE can be performed as an arrayed primer extension (APEX) reaction, in which primers for extension are fixed on a glass slide (Shumaker et al. 1996); polymermase and labeled ddNTPs are added to allow SBE reaction to proceed. As for the detection of OLA products, microarrays can be used to immobilize one of the ligation probes and allow OLA reaction to happen on the slide (Gerry et al. 1999).

Microbeads

Microbeads are small particles that can be linked to a probe. They are essentially a flexible form of a microarray, allowing flexibility in the assay design. SBE (Chen et al. 2000)and OLA (Iannone et al. 2000) can be performed in fashions analogous to mixed microarrays. Hybridization kinetics is believed to be better on microbeads than on traditional planar microarrays, as beads can freely move in solution. Individual bead identification depends on the bead properties. One class is differentially labeled with fluorophores, which can be sorted by FACs (Chen et al. 2000; Iannone et al. 2000). Beads barcoded with bound oligonucleotides can be coupled to a planar array for detection (Shen et al. 2005).

Solution phase:

Electrophoresis

To detect SBE products, capillary electrophoresis can be used to detect single base extension product from different fluorescently labeled dideoxynucleotides. Capillary

electrophoresis can be performed using channel capillaries as in a standard sequencing machine or on a microplate.

Gradient separation

One form of gradient separation is dynamic allelic-specific hybridization (DASH), in which a double-stranded DNA binding dye is used to monitor hybridization dynamics over a sweeping gradient of temperature, allowing more robust genotyping (Prince et al. 2001). An analogous idea using an electric field has also been reported (Sosnowski et al. 1997). Specialized instruments are required to set up these gradients.

Comments on SNP genotyping

The utilities of SNP genotyping in cancer research are wide. The number of approaches to genotype them seems to be bound only by human imagination. As mentioned, the choice of genotyping methods depends on the need of the experimenter. A whole-genome study will require a high-throughput format such as a microarray-based assay. The Affymetrix GeneChip has been the choice of most current whole-genome studies. Another highly multiplexed assay ideal for whole-genome studies is Illumina's GoldenGate assay, which combines microbeads and microarray on an array of arrays format (Table 1). Currently, a 100K GeneChip is available from Affymetrix and a 4.7K SNP linkage panel is available from Illumina (Matsuzaki et al. 2004; Murray et al. 2004). Capital investment required is the constraint. Other options such as electrophoresis based detection assays may provide the more cost-effective alternatives for smaller-scale studies. Other considerations in choosing an assay include the quantity and quality of the

available DNA. While genome amplification and complexity reduction methods have been developed for the current methods of SNP genotyping (Jordan et al. 2002; Barker et al. 2004), new detection means such as nanoparticles may make future assays sensitive enough to be performed on minute amount of samples without pre-processing. The rapidly evolving SNP genotyping technologies will hopefully lead to many cancer genomic discoveries in the years to come.

Digital Karyotyping

Aside from array CGH and SNP genotyping, another new genome analysis tool based on the genome sequence is digital karyotyping. The basic concept of digital karyotyping is similar to serial analysis of gene expression (SAGE) (Velculescu et al. 1995). As described by Wang et al. 2002, digital karyotyping involves isolation of 21bp short sequence tags from specific locations in the genome. Isolated tags are ligated into a concatenated form called 'ditags', which are then amplified by PCR en-masse and sequenced. Individual tags are digitally extracted from sequence data and matched on chromosomes. Tag densities along chromosomes are used to evaluate DNA content (Wang et al. 2002). Digital karyotyping is a powerful tool that has enabled the discovery of several specific cancer-associated gene amplifications, such as those of the homeobox gene *OTX2* in medulloblastoma (Boon et al. 2005), *Notch3* and the chromatin remodeling gene *Rsf-1/HBXAP* in ovarian cancer (Shih Ie et al. 2005; Park et al. 2006). The biggest limitation to perform digital karyotyping is cost.

Emerging Approaches to analyze the Sequenced Genome

In the previous section, a few emerging technologies: array CGH, SNP genotyping platforms, and digital karyotyping, have been described. These technologies provide means to analyze cancer genomes in a systematic fashion. Another set of opportunities involves approaching the available genome sequence in new ways. I will first describe the resequencing of certain genes in the human genome to identify cancer-related mutations and then the comparative genomic approach by studying animal cancers.

Genome resequencing

Mutations in signaling pathways involved in cell proliferation, cell death and cell differentiation are thought to be key in cancer. Resequencing of genes and gene families involved in these pathways has been pursued to systematically identify cancer-causing mutations. One of the first studies by Davis et al. coupled a heteroduplex-electrophoresis method with direct sequencing. They identified *BRAF* mutation in >60% of melanomas and at lower frequency in other cancers such as colorectoal cancer and non-small cell lung cancer (Davies et al. 2002). Subsequent exon resequencing experiments have focused on protein tyrosine phosphatases (*PTP*s), protein tyrosine kinases (*PTK*s) and phosphatidylinositol 3-kinases (*PI3K*s), identifying numerous cancer-specific mutations including *PTPRT, EGFR, ERBB2,* and *PIK3CA* (Paez et al. 2004; Samuels et al. 2004; Stephens et al. 2004; Wang et al. 2004). Resequencing the genome in a systematic manner is proving to be a fruitful approach.

Comparative Cancer Genomics

A comparative approach to study cancer is not a new idea. Many carcinogens in humans can similarly induce cancers in animals ---the first demonstration was tumor induction on rabbit ears from coal tar. Animal models of cancer can serve many research purposes: testing carcinogens, studying tumor biology, testing therapeutics etc.. The complete genome sequences of many model organisms are now available. Comparative analyses between genomes of different species have become feasible. In cancer research, comparative genomic studies are valuable in two major ways: 1) in validating animal models through an assessment of their degree of genetic resemblance to human disease and 2) in identifying genes and/or gene sets that are common to the model organism and human tumorigenesis.

Mouse cancer genomics

The genetic tractability of the mouse has made it an important animal model to study cancer (Van Dyke and Jacks 2002). A good mouse model should share phenotypic and genetic similarities to the human cancer it mimics (Hann and Balmain 2001). Genomics has aided the comparison of mouse models at both levels. Phenotypically, the use of global gene-expression as a validation tool between mouse and human cancers has been demonstrated (Sweet-Cordero et al. 2005). In addition, Sweet-Cordero et al. have shown that genome-wide gene expression data from controlled mouse experiments can help to filter molecular data from human samples. Genetically, a similar comparison can be performed using genomics tools that assess global genetic alterations. As in human cancers, chromosomal abnormalities have long been observed in tumors developed in

mice (Sasaki 1982; Liyanage et al. 1996). In the past few years, the sequencing of the mouse genome has demonstrated the high-degree of conservation between mouse and human genomes and spurred the development of high-resolution tools to characterize genetic lesions in mouse tumors.

The mouse genome

The first complete sequence of the mouse was published in 2002 (Waterston et al. 2002). Some observations by the authors are summarized below. The mouse genome is slightly smaller than the human genome (2.5Gb vs. 2.9Gb) but contains about the same number of genes (30,000), while 80% of the mouse genes have a single identifiable ortholog in human. At a gross chromosomal level, 75 million years of independent evolution has resulted in many large-scale rearrangements but local gene orders are mostly maintained. In fact, about 90% of the mouse and human genomes can be divided into regions of conserved of synteny (i.e. same thread), where local structure is intact. The total amount of ~350 conserved syntenic segments have been evolutionarily shuffled throughout the mouse and human genomes.

In thinking about comparative cancer genomics, the conservation of genes at 1:1 ratio suggests similar sets of genes likely control the same cellular processes in the two species. In addition, the partitioning of syntenic regions on different chromosomes provides a framework to assess the relative importance of each region in the other species.

Analyzing the mouse cancer genome

As in the old days of human cancer research, cytogenetics was the most accessible means to perform genomic analyses. Unlike human chromosomes, mouse chromosomes are acrocentric and similar in size. Karyotypic analysis in mouse cells were difficult until the development of chromosome painting techniques such as spectral karyotyping (Liyanage et al. 1996). In more recent years, mouse genome sequence availability has led to new tools for analysis. For copy number study, multiple array CGH platforms have been developed and the competition for better resolution has been fierce: starting from a custom-made BAC array covering the genome at 2-20Mb resolution (Hodgson et al. 2001), followed by BAC arrays with 1K probes (Cai et al. 2002), 2K probes (Snijders et al. 2005), 3K probes (Chung et al. 2004), and 19K probes (Li et al. 2004). The 19K array has clones spaced ~39kb throughout the genome (Li et al. 2004). Oligonucleotide CGH arrays have also been applied for mouse genome analysis: 20K array made by Agilent was reported to have a ~50kb genomic resolution (Brennan et al. 2004) while a non-commercially made 20K array has also been reported (van den Ijssel et al. 2005). Oligonucleotide arrays with> 40K probes are now available in the market. As for LOH mapping, several genome-wide SNP genotyping methods have been developed for the mouse (Petkov et al. 2004a; Owens et al. 2005; Moran et al. 2006). However, report that uses SNP in genome-wide LOH mapping in the mouse has not been made.

While new genomic tools for mouse tumor analysis are being refined or developed, a step back to summarize current data from mouse tumor genome studies would be appropriate.

Genetically engineered mouse cancer models enable the study of tumorigenesis in a controlled and reproducible fashion. Cooperating genetic lesions that enhance the tumorigenicity of the initiating mutant cells can be analyzed by array CGH. Several generalizations can be made from studying these models. First, large-scale chromosomal lesions appear to be predominant in telomerase active mice but the occasionally observed focal lesions have been helpful in pinpointing genetic regions important for tumorigenesis (Hodgson et al. 2001; Hackett et al. 2003). On the other hand, mice with dysfunctional telomeres exhibit a wider range of chromosomal abnormalities including more focal lesions (O'Hagan et al. 2002). Secondly, genetic background affects the lesions that are present (Hager et al. 2004). Thirdly, expression timing of the initiating mutation in a genetically engineered model can also influence the genetic alterations appeared (Hager et al. 2004). Finally, aside from identifying cancer genes, array CGH data can be used to classify tumors (O'Hagan et al. 2003). Some details leading to these sweeping statements are discussed below in the light of genetically engineered mouse models of various types cancers, including: 1) pancreatic islet carcinoma; 2) melanoma; 3) neuroblastoma, and 4) carcinomas including breast, colon, and skin tumors.

Pancreatic islet carcinoma model

The RIP-Tag mice express SV40 T antigens (Tag) under the control of the rat insulin promoter (RIP). In the initial pancreatic islet carcinoma study in RIP-Tag mice by Hodgson et al., most copy number alterations were observed to span large chromosomal areas but a focal lesion as small as ~3Mb could be detected on chromosome 16. That had allowed the authors to narrow down a previously known LOH region in the area that is

syntenic to chromosome 3q in human (Hodgson et al. 2001). In addition, new observations of other recurrent chromosomal copy number changes let the authors identify a few candidate oncogenes or tumor-suppressor genes (Hodgson et al. 2001), illustrating the potential value in performing genomic analysis on mouse tumors. A later CGH study on the islet tumors showed that RIP-Tag mice on FVB/N, C57Bl/6, and C4Heb/Fe backgrounds develop tumors with different copy number change spectra, suggesting the influence of genetic background (Hager et al. 2004). This is analogous to the varying susceptibility to different cancers within the human population. Another interesting finding in the islet cell cancer model was the timing effect of T-antigen expression on the copy number alterations seen (Hager et al. 2004). The authors proposed changes in tumor microenvironment at different time points can impose a different set of selection criteria on tumor cells.

Melanoma model

Melanomas develop in a *RAS*-induced $p19^{Arf}$-/- mouse model spontaneously but are accelerated by UV irradiation (Kannan et al. 2003). A use of array CGH data is tumor classification, as demonstrated in the classification of UV-induced vs. non-UV induced melanomas in these mice (O'Hagan et al. 2003).

Neuroblastoma model

Amplification of *MYCN* is frequently observed in human neuroblastoma (Brodeur et al. 1984). Neuroblastomas can be induced in mice by expressing human *MYCN* under a rat tyrosine hydroxylase (TH) promoter (Weiss et al. 1997). By CGH, several recurrent

whole chromosomal gains and losses appeared to cooperate with MYCN to drive tumorigenesis in this model (Hackett et al. 2003). More interestingly, Hackett et al. were able to identify a minimally gained region on mouse chromosome 11 by aligning recurrent focal gains that were observed. A syntenic comparison of the region allowed delineation of a frequently gained region on human chromosome 17q to 15Mb (Hackett et al. 2003).

Other carcinomas

Carcinomas are predominant in aging humans, but tumor spectrum in mice is skewed toward a high incidence of lymphomas and soft tissue sarcomas (Artandi et al. 2000). Mutating *p53* and *mTerc* (telomerase RNA component) not only shifts the mouse tumor spectrum to more human like but also leads to tumors with genetic aberrations more like those seen in human cancers; frequent aneuploidy, unbalanced translocations, amplifications and deletions are seen in tumors of these mice in the breast, colon, and lung (Artandi et al. 2000; O'Hagan et al. 2002). Significantly, genomic analysis by CGH suggested some of the recurrent copy number changes are syntenic to changes frequently seen in human cancers (O'Hagan et al. 2002).

What's now and what's next for mouse cancer genomics

Mouse models of cancer have become an integral part of cancer research for studying tumor biology and testing therapeutics. Using genomic analysis tools such as array CGH, the models can be validated for their degree of genetic resemblance to humans and be used to pinpoint the critical changes in human cancers. In addition, the availability of

whole mouse genome sequence has made whole-genome association studies feasible in mice, easing the discovery of disease loci.

Dog cancer genomics

Aside from the mouse, the value of the domestic dog in cancer research is becoming realized. To complete the discussion on comparative cancer genomics, I will briefly mention the role of dogs in cancer research.

The availability of many inbred strains has greatly facilitated mouse genetics. For dog, the extensive breeding of dogs through history has created many purebred strains. Genetic studies can be performed with less genetic background noise. In addition, many of these pure breeds show different susceptibility to diseases including cancer (Lindblad-Toh et al. 2005). As in mouse and humans, genetic association studies can be performed in dogs to identify genetic susceptibility loci.

The dog genome sequence is now complete (Lindblad-Toh et al. 2005). The dog and human genomes sequence are highly homologous; over 90% of the dog sequence lies in regions of conserved synteny with humans. As in the case between human and mouse, segments of synteny are distributed throughout each of the dog and human genomes during evolution, enabling a comparison of the relative role of individual syntenic segment in each species.

Like in human and mouse, recurrent chromosome aberrations have been observed in canine tumors (Dunn et al. 2000; Thomas et al. 2003; Milne et al. 2004). Characterization of genetic abnormalities may aid the discovery of orthologous cancer genes in humans. Recently, a 2-Mb resolution BAC microarray has been developed for CGH analysis of dog tumors (Thomas et al. 2005). The authors reported an osteosarcoma case that exhibited a wide range of abnormalities. Ongoing studies by the authors on a range of canine cancers including lymphoma, leukemia, osteosarcoma, and brain tumors were noted. In the years to come, a three-way comparison of tumor genomics between human, mouse, and dogs might yield interesting clues about the tumorigenesis process.

Outlook

Cancer genome sequencing

Genomic analysis is instrumental for the discovery of many underlying mutations in cancer. Some of the early discoveries of oncogenes have led to the development of a new class of target cancer drugs such as Gleevec and Herceptin. Evolving genomic technologies have provided the platforms further insights. It is worth mentioning that on December 13, 2005, the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI) launched a pilot project to build The Cancer Genome Atlas (TCGA). It is an initiative to apply genome analysis technologies, including large-scale genome sequencing, to study cancer. According to the mission statement, the goal of the pilot is to "assess the feasibility of a full-scale effort to

systematically explore the entire spectrum of genomic changes involved in human cancer" (NCI 2005). Future years in cancer genetics will likely be exciting.

Thesis Scope

The work presented in this thesis took a comparative approach to study cancer. Genomic studies were performed on mouse tumor models using two evolving tools: ROMA in characterizing copy number changes and genome-wide SNP genotyping in uncovering LOH regions. While the human ROMA platform has shown promise in delivering high-resolution data, the mouse version of ROMA needed to be tested for comparative genomic analysis purposes. Likewise, although the abundance of SNPs has made them valuable markers for LOH mapping in human tumors, the same concept had to be tested in mice. In proving the concept, a new protocol of SNP genotyping in mice was worked out. Chapters 2 and 3 describe the application of these two techniques in analyzing mouse models that mimic human lung cancer and retinoblastoma. In both mouse models, tumorigenesis was driven by mutations engineered in specific genes, *KRas* in the case of lung cancer and *Rb* and *p130* in the retinoblastoma model. Cooperating mutations that arose in the initiating mutant cells during tumorigenesis were examined in this study by the genomic tools described. Recurrent genetic changes could be detected in both models.

References

Albertson, D.G. and Pinkel, D. 2003. Genomic microarrays in human genetic disease and cancer. *Hum Mol Genet* **12 Spec No 2**: R145-152.

Artandi, S.E., Chang, S., Lee, S.L., Alson, S., Gottlieb, G.J., Chin, L., and DePinho, R.A. 2000. Telomere dysfunction promotes non-reciprocal translocations and epithelial cancers in mice. *Nature* **406**(6796): 641-645.

Balsara, B.R. and Testa, J.R. 2002. Chromosomal imbalances in human lung cancer. *Oncogene* **21**(45): 6877-6883.

Barker, D.L., Hansen, M.S., Faruqi, A.F., Giannola, D., Irsula, O.R., Lasken, R.S., Latterich, M., Makarov, V., Oliphant, A., Pinter, J.H., Shen, R., Sleptsova, I., Ziehler, W., and Lai, E. 2004. Two methods of whole-genome amplification enable accurate genotyping across a 2320-SNP linkage panel. *Genome Res* **14**(5): 901-907.

Barrett, M.T., Scheffer, A., Ben-Dor, A., Sampas, N., Lipson, D., Kincaid, R., Tsang, P., Curry, B., Baird, K., Meltzer, P.S., Yakhini, Z., Bruhn, L., and Laderman, S. 2004. Comparative genomic hybridization using oligonucleotide microarrays and total genomic DNA. *Proc Natl Acad Sci U S A* **101**(51): 17765-17770.

Ben-Avi, L., Durst, R., Shpitzen, S., Leitersdorf, E., and Meiner, V. 2004. Apolipoprotein E genotyping: accurate, simple, high throughput method using ABI Prism SNaPshot Multiplex System. *J Alzheimers Dis* **6**(5): 497-501.

Benavides, F., Conti, C.J., LaCava, M., Flores, M., Glasscock, E., Sternik, G., Gimenez-Conti, I.B., Johnston, D.A., Dunsford, H.A., Goldstein, L.S., and Rodriguez, L.V. 2003. Loss of heterozygosity analysis of mouse pulmonary adenomas induced by coal tar. *Environ Mol Mutagen* **41**(4): 300-308.

Beroukhim, R., Lin, M., Park, Y., Hao, K., Zhao, X., Garraway, L.A., Fox, E.A., Hochberg, E.P., Mellinghoff, I.K., Hofer, M.D., Descazeaud, A., Rubin, M.A., Meyerson, M., Wong, W.H., Sellers, W.R., and Li, C. 2006. Inferring loss-of-heterozygosity from unpaired tumors using high-density oligonucleotide SNP arrays. *PLoS Comput Biol* **2**(5): e41.

Bignell, G.R., Huang, J., Greshock, J., Watt, S., Butler, A., West, S., Grigorova, M., Jones, K.W., Wei, W., Stratton, M.R., Futreal, P.A., Weber, B., Shapero, M.H.,

and Wooster, R. 2004. High-resolution analysis of DNA copy number using oligonucleotide microarrays. *Genome Res* **14**(2): 287-295.

Boon, K., Eberhart, C.G., and Riggins, G.J. 2005. Genomic amplification of orthodenticle homologue 2 in medulloblastomas. *Cancer Res* **65**(3): 703-707.

Bray, M.S., Boerwinkle, E., and Doris, P.A. 2001. High-throughput multiplex SNP genotyping with MALDI-TOF mass spectrometry: practice, problems and promise. *Hum Mutat* **17**(4): 296-304.

Brennan, C., Zhang, Y., Leo, C., Feng, B., Cauwels, C., Aguirre, A.J., Kim, M., Protopopov, A., and Chin, L. 2004. High-resolution global profiling of genomic alterations with long oligonucleotide microarray. *Cancer Res* **64**(14): 4744-4748.

Brodeur, G.M., Seeger, R.C., Schwab, M., Varmus, H.E., and Bishop, J.M. 1984. Amplification of N-myc in untreated human neuroblastomas correlates with advanced disease stage. *Science* **224**(4653): 1121-1124.

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 Biotechnol* **20**(4): 393-396.

Cao, X., Eu, K.W., Kumarasinghe, M.P., Li, H.H., Loi, C., and Cheah, P.Y. 2006. Mapping of hereditary mixed polyposis syndrome (HMPS) to chromosome 10q23 by genomewide high-density single nucleotide polymorphism (SNP) scan and identification of BMPR1A loss of function. *J Med Genet* **43**(3): e13.

Carrasco, D.R., Tonon, G., Huang, Y., Zhang, Y., Sinha, R., Feng, B., Stewart, J.P., Zhan, F., Khatry, D., Protopopova, M., Protopopov, A., Sukhdeo, K., Hanamura, I., Stephens, O., Barlogie, B., Anderson, K.C., Chin, L., Shaughnessy, J.D., Jr., Brennan, C., and Depinho, R.A. 2006. High-resolution genomic profiles define distinct clinico-pathogenetic subgroups of multiple myeloma patients. *Cancer Cell* **9**(4): 313-325.

Carvalho, B., Ouwerkerk, E., Meijer, G.A., and Ylstra, B. 2004. High resolution microarray comparative genomic hybridisation analysis using spotted oligonucleotides. *J Clin Pathol* **57**(6): 644-646.

Chao, R.C., Pyzel, U., Fridlyand, J., Kuo, Y.M., Teel, L., Haaga, J., Borowsky, A., Horvai, A., Kogan, S.C., Bonifas, J., Huey, B., Jacks, T.E., Albertson, D.G., and

Shannon, K.M. 2005. Therapy-induced malignant neoplasms in Nf1 mutant mice. *Cancer Cell* **8**(4): 337-348.

Chen, D., Gallie, B.L., and Squire, J.A. 2001. Minimal regions of chromosomal imbalance in retinoblastoma detected by comparative genomic hybridization. *Cancer Genet Cytogenet* **129**(1): 57-63.

Chen, J., Iannone, M.A., Li, M.S., Taylor, J.D., Rivers, P., Nelsen, A.J., Slentz-Kesler, K.A., Roses, A., and Weiner, M.P. 2000. A microsphere-based assay for multiplexed single nucleotide polymorphism analysis using single base chain extension. *Genome Res* **10**(4): 549-557.

Chung, Y.J., Jonkers, J., Kitson, H., Fiegler, H., Humphray, S., Scott, C., Hunt, S., Yu, Y., Nishijima, I., Velds, A., Holstege, H., Carter, N., and Bradley, A. 2004. A whole-genome mouse BAC microarray with 1-Mb resolution for analysis of DNA copy number changes by array comparative genomic hybridization. *Genome Res* **14**(1): 188-196.

Davies, H., Bignell, G.R., Cox, C., Stephens, P., Edkins, S., Clegg, S., Teague, J., Woffendin, H., Garnett, M.J., Bottomley, W., Davis, N., Dicks, E., Ewing, R., Floyd, Y., Gray, K., Hall, S., Hawes, R., Hughes, J., Kosmidou, V., Menzies, A., Mould, C., Parker, A., Stevens, C., Watt, S., Hooper, S., Wilson, R., Jayatilake, H., Gusterson, B.A., Cooper, C., Shipley, J., Hargrave, D., Pritchard-Jones, K., Maitland, N., Chenevix-Trench, G., Riggins, G.J., Bigner, D.D., Palmieri, G., Cossu, A., Flanagan, A., Nicholson, A., Ho, J.W., Leung, S.Y., Yuen, S.T., Weber, B.L., Seigler, H.F., Darrow, T.L., Paterson, H., Marais, R., Marshall, C.J., Wooster, R., Stratton, M.R., and Futreal, P.A. 2002. Mutations of the BRAF gene in human cancer. *Nature* **417**(6892): 949-954.

Dhami, P., Coffey, A.J., Abbs, S., Vermeesch, J.R., Dumanski, J.P., Woodward, K.J., Andrews, R.M., Langford, C., and Vetrie, D. 2005. Exon array CGH: detection of copy-number changes at the resolution of individual exons in the human genome. *Am J Hum Genet* **76**(5): 750-762.

Dietrich, W.F., Radany, E.H., Smith, J.S., Bishop, J.M., Hanahan, D., and Lander, E.S. 1994. Genome-wide search for loss of heterozygosity in transgenic mouse tumors reveals candidate tumor suppressor genes on chromosomes 9 and 16. *Proc Natl Acad Sci U S A* **91**(20): 9451-9455.

Dunn, K.A., Thomas, R., Binns, M.M., and Breen, M. 2000. Comparative genomic hybridization (CGH) in dogs--application to the study of a canine glial tumour cell line. *Vet J* **160**(1): 77-82.

Edwards, J.R., Ruparel, H., and Ju, J. 2005. Mass-spectrometry DNA sequencing. *Mutat Res* **573**(1-2): 3-12.

Einarsdottir, K., Humphreys, K., Bonnard, C., Palmgren, J., Iles, M.M., Sjolander, A., Li, Y., Chia, K.S., Liu, E.T., Hall, P., Liu, J., and Wedren, S. 2006. Linkage Disequilibrium Mapping of CHEK2: Common Variation and Breast Cancer Risk. *PLoS Med* **3**(6): e168.

Ellis, N.A., Kirchhoff, T., Mitra, N., Ye, T.Z., Chuai, S., Huang, H., Nafa, K., Norton, L., Neuhausen, S., Gordon, D., Struewing, J.P., Narod, S., and Offit, K. 2006. Localization of breast cancer susceptibility loci by genome-wide SNP linkage disequilibrium mapping. *Genet Epidemiol* **30**(1): 48-61.

Engle, L.J., Simpson, C.L., and Landers, J.E. 2006. Using high-throughput SNP technologies to study cancer. *Oncogene* **25**(11): 1594-1601.

Finney, R.E. and Bishop, J.M. 1993. Predisposition to neoplastic transformation caused by gene replacement of H-ras1. *Science* **260**(5113): 1524-1527.

Frazer, K.A., Wade, C.M., Hinds, D.A., Patil, N., Cox, D.R., and Daly, M.J. 2004. Segmental phylogenetic relationships of inbred mouse strains revealed by fine-scale analysis of sequence variation across 4.6 mb of mouse genome. *Genome Res* **14**(8): 1493-1500.

Friedhoff, P., Hahn, M., Wolfes, H., and Pingoud, A. 1993. Quantitative polymerase chain reaction with oligodeoxynucleotide ligation assay/enzyme-linked immunosorbent assay detection. *Anal Biochem* **215**(1): 9-16.

Fujiwara, T., Bandi, M., Nitta, M., Ivanova, E.V., Bronson, R.T., and Pellman, D. 2005. Cytokinesis failure generating tetraploids promotes tumorigenesis in p53-null cells. *Nature* **437**(7061): 1043-1047.

Futreal, P.A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., and Stratton, M.R. 2004. A census of human cancer genes. *Nat Rev Cancer* **4**(3): 177-183.

Garnis, C., Lockwood, W.W., Vucic, E., Ge, Y., Girard, L., Minna, J.D., Gazdar, A.F., Lam, S., Macaulay, C., and Lam, W.L. 2006. High resolution analysis of non-small cell lung cancer cell lines by whole genome tiling path array CGH. *Int J Cancer* **118**(6): 1556-1564.

Germer, S., Holland, M.J., and Higuchi, R. 2000. High-throughput SNP allele-frequency determination in pooled DNA samples by kinetic PCR. *Genome Res* **10**(2): 258-266.

Gerry, N.P., Witowski, N.E., Day, J., Hammer, R.P., Barany, G., and Barany, F. 1999. Universal DNA microarray method for multiplex detection of low abundance point mutations. *J Mol Biol* **292**(2): 251-262.

Gilles, P.N., Wu, D.J., Foster, C.B., Dillon, P.J., and Chanock, S.J. 1999. Single nucleotide polymorphic discrimination by an electronic dot blot assay on semiconductor microchips. *Nat Biotechnol* **17**(4): 365-370.

Grupe, A., Germer, S., Usuka, J., Aud, D., Belknap, J.K., Klein, R.F., Ahluwalia, M.K., Higuchi, R., and Peltz, G. 2001. In silico mapping of complex disease-related traits in mice. *Science* **292**(5523): 1915-1918.

Hackett, C.S., Hodgson, J.G., Law, M.E., Fridlyand, J., Osoegawa, K., de Jong, P.J., Nowak, N.J., Pinkel, D., Albertson, D.G., Jain, A., Jenkins, R., Gray, J.W., and Weiss, W.A. 2003. Genome-wide array CGH analysis of murine neuroblastoma reveals distinct genomic aberrations which parallel those in human tumors. *Cancer Res* **63**(17): 5266-5273.

Hager, J.H., Hodgson, J.G., Fridlyand, J., Hariono, S., Gray, J.W., and Hanahan, D. 2004. Oncogene expression and genetic background influence the frequency of DNA copy number abnormalities in mouse pancreatic islet cell carcinomas. *Cancer Res* **64**(7): 2406-2410.

Hamaguchi, M., Meth, J.L., von Klitzing, C., Wei, W., Esposito, D., Rodgers, L., Walsh, T., Welcsh, P., King, M.C., and Wigler, M.H. 2002. DBC2, a candidate for a tumor suppressor gene involved in breast cancer. *Proc Natl Acad Sci U S A* **99**(21): 13647-13652.

Han, M., Gao, X., Su, J.Z., and Nie, S. 2001. Quantum-dot-tagged microbeads for multiplexed optical coding of biomolecules. *Nat Biotechnol* **19**(7): 631-635.

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

Hann, B. and Balmain, A. 2001. Building 'validated' mouse models of human cancer. *Curr Opin Cell Biol* **13**(6): 778-784.

Hemann, M.T., Fridman, J.S., Zilfou, J.T., Hernando, E., Paddison, P.J., Cordon-Cardo, C., Hannon, G.J., and Lowe, S.W. 2003. An epi-allelic series of p53 hypomorphs created by stable RNAi produces distinct tumor phenotypes in vivo. *Nat Genet* **33**(3): 396-400.

Herzog, C.R., Devereux, T.R., Pittman, B., and You, M. 2002. Carcinogenic induction directs the selection of allelic losses in mouse lung tumorigenesis. *Cancer Res* **62**(22): 6424-6429.

Hingorani, S.R., Wang, L., Multani, A.S., Combs, C., Deramaudt, T.B., Hruban, R.H., Rustgi, A.K., Chang, S., and Tuveson, D.A. 2005. Trp53R172H and KrasG12D cooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice. *Cancer Cell* **7**(5): 469-483.

Hodgson, G., Hager, J.H., Volik, S., Hariono, S., Wernick, M., Moore, D., Nowak, N., 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**(4): 459-464.

Hoglund, M., Gisselsson, D., Hansen, G.B., and Mitelman, F. 2004. Statistical dissection of cytogenetic patterns in lung cancer reveals multiple modes of karyotypic evolution independent of histological classification. *Cancer Genet Cytogenet* **154**(2): 99-109.

Huang, J., Wei, W., Zhang, J., Liu, G., Bignell, G.R., Stratton, M.R., Futreal, P.A., Wooster, R., Jones, K.W., and Shapero, M.H. 2004. Whole genome DNA copy number changes identified by high density oligonucleotide arrays. *Hum Genomics* **1**(4): 287-299.

Iannone, M.A., Taylor, J.D., Chen, J., Li, M.S., Rivers, P., Slentz-Kesler, K.A., and Weiner, M.P. 2000. Multiplexed single nucleotide polymorphism genotyping by oligonucleotide ligation and flow cytometry. *Cytometry* **39**(2): 131-140.

Ishkanian, A.S., Malloff, C.A., Watson, S.K., DeLeeuw, R.J., Chi, B., Coe, B.P., Snijders, A., Albertson, D.G., Pinkel, D., Marra, M.A., Ling, V., MacAulay, C.,

and Lam, W.L. 2004. A tiling resolution DNA microarray with complete coverage of the human genome. *Nat Genet* **36**(3): 299-303.

Jackson, E.L., Olive, K.P., Tuveson, D.A., Bronson, R., Crowley, D., Brown, M., and Jacks, T. 2005. The differential effects of mutant p53 alleles on advanced murine lung cancer. *Cancer Res* **65**(22): 10280-10288.

Janne, P.A., Li, C., Zhao, X., Girard, L., Chen, T.H., Minna, J., Christiani, D.C., Johnson, B.E., and Meyerson, M. 2004. High-resolution single-nucleotide polymorphism array and clustering analysis of loss of heterozygosity in human lung cancer cell lines. *Oncogene* **23**(15): 2716-2726.

Jobanputra, V., Sebat, J., Troge, J., Chung, W., Anyane-Yeboa, K., Wigler, M., and Warburton, D. 2005. Application of ROMA (representational oligonucleotide microarray analysis) to patients with cytogenetic rearrangements. *Genet Med* **7**(2): 111-118.

Johnson, L., Mercer, K., Greenbaum, D., Bronson, R.T., Crowley, D., Tuveson, D.A., and Jacks, T. 2001. Somatic activation of the K-ras oncogene causes early onset lung cancer in mice. *Nature* **410**(6832): 1111-1116.

Jordan, B., Charest, A., Dowd, J.F., Blumenstiel, J.P., Yeh Rf, R.F., Osman, A., Housman, D.E., and Landers, J.E. 2002. Genome complexity reduction for SNP genotyping analysis. *Proc Natl Acad Sci U S A* **99**(5): 2942-2947.

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**(5083): 818-821.

Kannan, K., Sharpless, N.E., Xu, J., O'Hagan, R.C., Bosenberg, M., and Chin, L. 2003. Components of the Rb pathway are critical targets of UV mutagenesis in a murine melanoma model. *Proc Natl Acad Sci U S A* **100**(3): 1221-1225.

Kim, T.M., Yim, S.H., Lee, J.S., Kwon, M.S., Ryu, J.W., Kang, H.M., Fiegler, H., Carter, N.P., and Chung, Y.J. 2005. Genome-wide screening of genomic alterations and their clinicopathologic implications in non-small cell lung cancers. *Clin Cancer Res* **11**(23): 8235-8242.

Knudson, A.G., Jr. 1971. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* **68**(4): 820-823.

Kuimelis, R.G., Livak, K.J., Mullah, B., and Andrus, A. 1997. Structural analogues of TaqMan probes for real-time quantitative PCR. *Nucleic Acids Symp Ser*(37): 255-256.

Kuukasjarvi, T., Tanner, M., Pennanen, S., Karhu, R., Visakorpi, T., and Isola, J. 1997. Optimizing DOP-PCR for universal amplification of small DNA samples in comparative genomic hybridization. *Genes Chromosomes Cancer* **18**(2): 94-101.

Lage, J.M., Leamon, J.H., Pejovic, T., Hamann, S., Lacey, M., Dillon, D., Segraves, R., Vossbrinck, B., Gonzalez, A., Pinkel, D., Albertson, D.G., Costa, J., and Lizardi, P.M. 2003. Whole genome analysis of genetic alterations in small DNA samples using hyperbranched strand displacement amplification and array-CGH. *Genome Res* **13**(2): 294-307.

Landegren, U., Kaiser, R., Sanders, J., and Hood, L. 1988. A ligase-mediated gene detection technique. *Science* **241**(4869): 1077-1080.

Lander, E.S. Linton, L.M. Birren, B. Nusbaum, C. Zody, M.C. Baldwin, J. Devon, K. Dewar, K. Doyle, M. FitzHugh, W. Funke, R. Gage, D. Harris, K. Heaford, A. Howland, J. Kann, L. Lehoczky, J. LeVine, R. McEwan, P. McKernan, K. Meldrim, J. Mesirov, J.P. Miranda, C. Morris, W. Naylor, J. Raymond, C. Rosetti, M. Santos, R. Sheridan, A. Sougnez, C. Stange-Thomann, N. Stojanovic, N. Subramanian, A. Wyman, D. Rogers, J. Sulston, J. Ainscough, R. Beck, S. Bentley, D. Burton, J. Clee, C. Carter, N. Coulson, A. Deadman, R. Deloukas, P. Dunham, A. Dunham, I. Durbin, R. French, L. Grafham, D. Gregory, S. Hubbard, T. Humphray, S. Hunt, A. Jones, M. Lloyd, C. McMurray, A. Matthews, L. Mercer, S. Milne, S. Mullikin, J.C. Mungall, A. Plumb, R. Ross, M. Shownkeen, R. Sims, S. Waterston, R.H. Wilson, R.K. Hillier, L.W. McPherson, J.D. Marra, M.A. Mardis, E.R. Fulton, L.A. Chinwalla, A.T. Pepin, K.H. Gish, W.R. Chissoe, S.L. Wendl, M.C. Delehaunty, K.D. Miner, T.L. Delehaunty, A. Kramer, J.B. Cook, L.L. Fulton, R.S. Johnson, D.L. Minx, P.J. Clifton, S.W. Hawkins, T. Branscomb, E. Predki, P. Richardson, P. Wenning, S. Slezak, T. Doggett, N. Cheng, J.F. Olsen, A. Lucas, S. Elkin, C. Uberbacher, E. Frazier, M. Gibbs, R.A. Muzny, D.M. Scherer, S.E. Bouck, J.B. Sodergren, E.J. Worley, K.C. Rives, C.M. Gorrell, J.H. Metzker, M.L. Naylor, S.L. Kucherlapati, R.S. Nelson, D.L. Weinstock, G.M. Sakaki, Y. Fujiyama, A. Hattori, M. Yada, T. Toyoda, A. Itoh, T. Kawagoe, C. Watanabe, H. Totoki, Y. Taylor, T. Weissenbach, J. Heilig, R. Saurin, W. Artiguenave, F. Brottier, P. Bruls, T. Pelletier, E. Robert, C. Wincker, P. Smith, D.R. Doucette-Stamm, L. Rubenfield, M. Weinstock, K. Lee, H.M. Dubois, J. Rosenthal, A. Platzer, M. Nyakatura, G. Taudien, S. Rump, A. Yang, H. Yu, J. Wang, J. Huang, G. Gu, J. Hood, L. Rowen, L. Madan, A. Qin, S. Davis, R.W. Federspiel, N.A. Abola, A.P. Proctor, M.J. Myers, R.M. Schmutz, J. Dickson, M. Grimwood, J. Cox, D.R. Olson, M.V. Kaul, R. Raymond, C. Shimizu, N. Kawasaki, K. Minoshima, S. Evans, G.A. Athanasiou, M. Schultz, R.

Roe, B.A. Chen, F. Pan, H. Ramser, J. Lehrach, H. Reinhardt, R. McCombie, W.R. de la Bastide, M. Dedhia, N. Blocker, H. Hornischer, K. Nordsiek, G. Agarwala, R. Aravind, L. Bailey, J.A. Bateman, A. Batzoglou, S. Birney, E. Bork, P. Brown, D.G. Burge, C.B. Cerutti, L. Chen, H.C. Church, D. Clamp, M. Copley, R.R. Doerks, T. Eddy, S.R. Eichler, E.E. Furey, T.S. Galagan, J. Gilbert, J.G. Harmon, C. Hayashizaki, Y. Haussler, D. Hermjakob, H. Hokamp, K. Jang, W. Johnson, L.S. Jones, T.A. Kasif, S. Kaspryzk, A. Kennedy, S. Kent, W.J. Kitts, P. Koonin, E.V. Korf, I. Kulp, D. Lancet, D. Lowe, T.M. McLysaght, A. Mikkelsen, T. Moran, J.V. Mulder, N. Pollara, V.J. Ponting, C.P. Schuler, G. Schultz, J. Slater, G. Smit, A.F. Stupka, E. Szustakowski, J. Thierry-Mieg, D. Thierry-Mieg, J. Wagner, L. Wallis, J. Wheeler, R. Williams, A. Wolf, Y.I. Wolfe, K.H. Yang, S.P. Yeh, R.F. Collins, F. Guyer, M.S. Peterson, J. Felsenfeld, A. Wetterstrand, K.A. Patrinos, A. Morgan, M.J. de Jong, P. Catanese, J.J. Osoegawa, K. Shizuya, H. Choi, S. and Chen, Y.J. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**(6822): 860-921.

Langaee, T. and Ronaghi, M. 2005. Genetic variation analyses by Pyrosequencing. *Mutat Res* **573**(1-2): 96-102.

Lanni, J.S. and Jacks, T. 1998. Characterization of the p53-dependent postmitotic checkpoint following spindle disruption. *Mol Cell Biol* **18**(2): 1055-1064.

Larson, G.P., Ding, Y., Cheng, L.S., Lundberg, C., Gagalang, V., Rivas, G., Geller, L., Weitzel, J., MacDonald, D., Archambeau, J., Slater, J., Neuberg, D., Daly, M.B., Angel, I., Benson, A.B., 3rd, Smith, K., Kirkwood, J.M., O'Dwyer, P.J., Raskay, B., Sutphen, R., Drew, R., Stewart, J.A., Werndli, J., Johnson, D., Ruckdeschel, J.C., Elston, R.C., and Krontiris, T.G. 2005. Genetic linkage of prostate cancer risk to the chromosome 3 region bearing FHIT. *Cancer Res* **65**(3): 805-814.

Li, J., Jiang, T., Mao, J.H., Balmain, A., Peterson, L., Harris, C., Rao, P.H., Havlak, P., Gibbs, R., and Cai, W.W. 2004. Genomic segmental polymorphisms in inbred mouse strains. *Nat Genet* **36**(9): 952-954.

Li, J., Yen, C., Liaw, D., Podsypanina, K., Bose, S., Wang, S.I., Puc, J., Miliaresis, C., Rodgers, L., McCombie, R., Bigner, S.H., Giovanella, B.C., Ittmann, M., Tycko, B., Hibshoosh, H., Wigler, M.H., and Parsons, R. 1997. PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. *Science* **275**(5308): 1943-1947.

Lillington, D.M., Kingston, J.E., Coen, P.G., Price, E., Hungerford, J., Domizio, P., Young, B.D., and Onadim, Z. 2003. Comparative genomic hybridization of 49 primary retinoblastoma tumors identifies chromosomal regions associated with

histopathology, progression, and patient outcome. *Genes Chromosomes Cancer* **36**(2): 121-128.

Lindblad-Toh, K., Tanenbaum, D.M., Daly, M.J., Winchester, E., Lui, W.O., Villapakkam, A., Stanton, S.E., Larsson, C., Hudson, T.J., Johnson, B.E., Lander, E.S., and Meyerson, M. 2000a. Loss-of-heterozygosity analysis of small-cell lung carcinomas using single-nucleotide polymorphism arrays. *Nat Biotechnol* **18**(9): 1001-1005.

Lindblad-Toh, K. Wade, C.M. Mikkelsen, T.S. Karlsson, E.K. Jaffe, D.B. Kamal, M. Clamp, M. Chang, J.L. Kulbokas, E.J., 3rd Zody, M.C. Mauceli, E. Xie, X. Breen, M. Wayne, R.K. Ostrander, E.A. Ponting, C.P. Galibert, F. Smith, D.R. DeJong, P.J. Kirkness, E. Alvarez, P. Biagi, T. Brockman, W. Butler, J. Chin, C.W. Cook, A. Cuff, J. Daly, M.J. DeCaprio, D. Gnerre, S. Grabherr, M. Kellis, M. Kleber, M. Bardeleben, C. Goodstadt, L. Heger, A. Hitte, C. Kim, L. Koepfli, K.P. Parker, H.G. Pollinger, J.P. Searle, S.M. Sutter, N.B. Thomas, R. Webber, C. Baldwin, J. Abebe, A. Abouelleil, A. Aftuck, L. Ait-Zahra, M. Aldredge, T. Allen, N. An, P. Anderson, S. Antoine, C. Arachchi, H. Aslam, A. Ayotte, L. Bachantsang, P. Barry, A. Bayul, T. Benamara, M. Berlin, A. Bessette, D. Blitshteyn, B. Bloom, T. Blye, J. Boguslavskiy, L. Bonnet, C. Boukhgalter, B. Brown, A. Cahill, P. Calixte, N. Camarata, J. Cheshatsang, Y. Chu, J. Citroen, M. Collymore, A. Cooke, P. Dawoe, T. Daza, R. Decktor, K. DeGray, S. Dhargay, N. Dooley, K. Dooley, K. Dorje, P. Dorjee, K. Dorris, L. Duffey, N. Dupes, A. Egbiremolen, O. Elong, R. Falk, J. Farina, A. Faro, S. Ferguson, D. Ferreira, P. Fisher, S. FitzGerald, M. Foley, K. Foley, C. Franke, A. Friedrich, D. Gage, D. Garber, M. Gearin, G. Giannoukos, G. Goode, T. Goyette, A. Graham, J. Grandbois, E. Gyaltsen, K. Hafez, N. Hagopian, D. Hagos, B. Hall, J. Healy, C. Hegarty, R. Honan, T. Horn, A. Houde, N. Hughes, L. Hunnicutt, L. Husby, M. Jester, B. Jones, C. Kamat, A. Kanga, B. Kells, C. Khazanovich, D. Kieu, A.C. Kisner, P. Kumar, M. Lance, K. Landers, T. Lara, M. Lee, W. Leger, J.P. Lennon, N. Leuper, L. LeVine, S. Liu, J. Liu, X. Lokyitsang, Y. Lokyitsang, T. Lui, A. Macdonald, J. Major, J. Marabella, R. Maru, K. Matthews, C. McDonough, S. Mehta, T. Meldrim, J. Melnikov, A. Meneus, L. Mihalev, A. Mihova, T. Miller, K. Mittelman, R. Mlenga, V. Mulrain, L. Munson, G. Navidi, A. Naylor, J. Nguyen, T. Nguyen, N. Nguyen, C. Nguyen, T. Nicol, R. Norbu, N. Norbu, C. Novod, N. Nyima, T. Olandt, P. O'Neill, B. O'Neill, K. Osman, S. Oyono, L. Patti, C. Perrin, D. Phunkhang, P. Pierre, F. Priest, M. Rachupka, A. Raghuraman, S. Rameau, R. Ray, V. Raymond, C. Rege, F. Rise, C. Rogers, J. Rogov, P. Sahalie, J. Settipalli, S. Sharpe, T. Shea, T. Sheehan, M. Sherpa, N. Shi, J. Shih, D. Sloan, J. Smith, C. Sparrow, T. Stalker, J. Stange-Thomann, N. Stavropoulos, S. Stone, C. Stone, S. Sykes, S. Tchuinga, P. Tenzing, P. Tesfaye, S. Thoulutsang, D. Thoulutsang, Y. Topham, K. Topping, I. Tsamla, T. Vassiliev, H. Venkataraman, V. Vo, A. Wangchuk, T. Wangdi, T. Weiand, M. Wilkinson, J. Wilson, A. Yadav, S. Yang, S. Yang, X. Young, G. Yu, Q. Zainoun, J. Zembek,

L. Zimmer, A. and Lander, E.S. 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**(7069): 803-819.

Lindblad-Toh, K., Winchester, E., Daly, M.J., Wang, D.G., Hirschhorn, J.N., Laviolette, J.P., Ardlie, K., Reich, D.E., Robinson, E., Sklar, P., Shah, N., Thomas, D., Fan, J.B., Gingeras, T., Warrington, J., Patil, N., Hudson, T.J., and Lander, E.S. 2000b. Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nat Genet* **24**(4): 381-386.

Lisitsyn, N. and Wigler, M. 1995. Representational difference analysis in detection of genetic lesions in cancer. *Methods Enzymol* **254**: 291-304.

Lisitsyn, N.A., Lisitsina, N.M., Dalbagni, G., Barker, P., Sanchez, C.A., Gnarra, J., Linehan, W.M., Reid, B.J., and Wigler, M.H. 1995. Comparative genomic analysis of tumors: detection of DNA losses and amplification. *Proc Natl Acad Sci U S A* **92**(1): 151-155.

Livak, K.J. 1999. Allelic discrimination using fluorogenic probes and the 5' nuclease assay. *Genet Anal* **14**(5-6): 143-149.

Livak, K.J., Flood, S.J., Marmaro, J., Giusti, W., and Deetz, K. 1995. Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *PCR Methods Appl* **4**(6): 357-362.

Liyanage, M., Coleman, A., du Manoir, S., Veldman, T., McCormack, S., Dickson, R.B., Barlow, C., Wynshaw-Boris, A., Janz, S., Wienberg, J., Ferguson-Smith, M.A., Schrock, E., and Ried, T. 1996. Multicolour spectral karyotyping of mouse chromosomes. *Nat Genet* **14**(3): 312-315.

Lucito, R., Healy, J., Alexander, J., Reiner, A., Esposito, D., Chi, M., Rodgers, L., Brady, A., Sebat, J., Troge, J., West, J.A., Rostan, S., Nguyen, K.C., Powers, S., Ye, K.Q., Olshen, A., Venkatraman, E., Norton, L., and Wigler, M. 2003. Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation. *Genome Res* **13**(10): 2291-2305.

Lucito, R., Nakimura, M., West, J.A., Han, Y., Chin, K., Jensen, K., McCombie, R., Gray, J.W., and Wigler, M. 1998. Genetic analysis using genomic representations. *Proc Natl Acad Sci U S A* **95**(8): 4487-4492.

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**(11): 1726-1736.

MacPherson, D., Conkrite, K., Tam, M., Mukai, S., Mu, D., and Jacks, T. submitted. Murine bilateral retinoblastoma exhibiting rapid onset, metastatic progression and N-myc gene amplification.

Mairal, A., Pinglier, E., Gilbert, E., Peter, M., Validire, P., Desjardins, L., Doz, F., Aurias, A., and Couturier, J. 2000. Detection of chromosome imbalances in retinoblastoma by parallel karyotype and CGH analyses. *Genes Chromosomes Cancer* **28**(4): 370-379.

Makridakis, N.M. and Reichardt, J.K. 2001. Multiplex automated primer extension analysis: simultaneous genotyping of several polymorphisms. *Biotechniques* **31**(6): 1374-1380.

Manchester, K.L. 1995. Theodor Boveri and the origin of malignant tumours. *Trends Cell Biol* **5**(10): 384-387.

Manenti, G., Galbiati, F., Gianni-Barrera, R., Pettinicchio, A., Acevedo, A., and Dragani, T.A. 2004. Haplotype sharing suggests that a genomic segment containing six genes accounts for the pulmonary adenoma susceptibility 1 (Pas1) locus activity in mice. *Oncogene* **23**(25): 4495-4504.

Matsuzaki, H., Dong, S., Loi, H., Di, X., Liu, G., Hubbell, E., Law, J., Berntsen, T., Chadha, M., Hui, H., Yang, G., Kennedy, G.C., Webster, T.A., Cawley, S., Walsh, P.S., Jones, K.W., Fodor, S.P., and Mei, R. 2004. Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. *Nat Methods* **1**(2): 109-111.

Mei, R., Galipeau, P.C., Prass, C., Berno, A., Ghandour, G., Patil, N., Wolff, R.K., Chee, M.S., Reid, B.J., and Lockhart, D.J. 2000. Genome-wide detection of allelic imbalance using human SNPs and high-density DNA arrays. *Genome Res* **10**(8): 1126-1137.

Meraldi, P., Honda, R., and Nigg, E.A. 2002. Aurora-A overexpression reveals tetraploidization as a major route to centrosome amplification in p53-/- cells. *Embo J* **21**(4): 483-492.

Milne, B.S., Hoather, T., O'Brien, P.C., Yang, F., Ferguson-Smith, M.A., Dobson, J., and Sargan, D. 2004. Karyotype of canine soft tissue sarcomas: a multi-colour, multi-species approach to canine chromosome painting. *Chromosome Res* **12**(8): 825-835.

Mitelman, F. 2000. Recurrent chromosome aberrations in cancer. *Mutat Res* **462**(2-3): 247-253.

Mitelman, F., Johansson, B., and Mertens, F. 2006. Mitelman Database of Chromosome Aberrations in Cancer. In. http://cgap.nci.nih.gov/Chromosomes/Mitelman.

Mitra, N., Ye, T.Z., Smith, A., Chuai, S., Kirchhoff, T., Peterlongo, P., Nafa, K., Phillips, M.S., Offit, K., and Ellis, N.A. 2004. Localization of cancer susceptibility genes by genome-wide single-nucleotide polymorphism linkage-disequilibrium mapping. *Cancer Res* **64**(21): 8116-8125.

Moran, J.L., Bolton, A.D., Tran, P.V., Brown, A., Dwyer, N.D., Manning, D.K., Bjork, B.C., Li, C., Montgomery, K., Siepka, S.M., Vitaterna, M.H., Takahashi, J.S., Wiltshire, T., Kwiatkowski, D.J., Kucherlapati, R., and Beier, D.R. 2006. Utilization of a whole genome SNP panel for efficient genetic mapping in the mouse. *Genome Res* **16**(3): 436-440.

Mu, D., Chen, L., Zhang, X., See, L.H., Koch, C.M., Yen, C., Tong, J.J., Spiegel, L., Nguyen, K.C., Servoss, A., Peng, Y., Pei, L., Marks, J.R., Lowe, S., Hoey, T., Jan, L.Y., McCombie, W.R., Wigler, M.H., and Powers, S. 2003. Genomic amplification and oncogenic properties of the KCNK9 potassium channel gene. *Cancer Cell* **3**(3): 297-302.

Murray, S.S., Oliphant, A., Shen, R., McBride, C., Steeke, R.J., Shannon, S.G., Rubano, T., Kermani, B.G., Fan, J.B., Chee, M.S., and Hansen, M.S. 2004. A highly informative SNP linkage panel for human genetic studies. *Nat Methods* **1**(2): 113-117.

NCI. 2005. The Cancer Genome Atlas. *http://cancergenomenihgov*

.

Nikiforov, T.T., Rendle, R.B., Goelet, P., Rogers, Y.H., Kotewicz, M.L., Anderson, S., Trainor, G.L., and Knapp, M.R. 1994. Genetic Bit Analysis: a solid phase method for typing single nucleotide polymorphisms. *Nucleic Acids Res* **22**(20): 4167-4175.

Norton, N., Williams, N.M., Williams, H.J., Spurlock, G., Kirov, G., Morris, D.W., Hoogendoorn, B., Owen, M.J., and O'Donovan, M.C. 2002. Universal, robust, highly quantitative SNP allele frequency measurement in DNA pools. *Hum Genet* **110**(5): 471-478.

Nowell, P.C. 1976. The clonal evolution of tumor cell populations. *Science* **194**(4260): 23-28.

Nowell, P.C. and Hungerford, D.A. 1960. A minute chromosome in human chronic granulocytic leukemia. *Science* **132**: 1497.

O'Hagan, R.C., Brennan, C.W., Strahs, A., Zhang, X., Kannan, K., Donovan, M., Cauwels, C., Sharpless, N.E., Wong, W.H., and Chin, L. 2003. Array comparative genome hybridization for tumor classification and gene discovery in mouse models of malignant melanoma. *Cancer Res* **63**(17): 5352-5356.

O'Hagan, R.C., Chang, S., Maser, R.S., Mohan, R., Artandi, S.E., Chin, L., and DePinho, R.A. 2002. Telomere dysfunction provokes regional amplification and deletion in cancer genomes. *Cancer Cell* **2**(2): 149-155.

Olivier, M. 2005. The Invader assay for SNP genotyping. *Mutat Res* **573**(1-2): 103-110.

Orum, H., Jakobsen, M.H., Koch, T., Vuust, J., and Borre, M.B. 1999. Detection of the factor V Leiden mutation by direct allele-specific hybridization of PCR amplicons to photoimmobilized locked nucleic acids. *Clin Chem* **45**(11): 1898-1905.

Owens, S.E., Broman, K.W., Wiltshire, T., Elmore, J.B., Bradley, K.M., Smith, J.R., and Southard-Smith, E.M. 2005. Genome-wide linkage identifies novel modifier loci of aganglionosis in the Sox10Dom model of Hirschsprung disease. *Hum Mol Genet* **14**(11): 1549-1558.

Paez, J.G., Janne, P.A., Lee, J.C., Tracy, S., Greulich, H., Gabriel, S., Herman, P., Kaye, F.J., Lindeman, N., Boggon, T.J., Naoki, K., Sasaki, H., Fujii, Y., Eck, M.J., Sellers, W.R., Johnson, B.E., and Meyerson, M. 2004. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* **304**(5676): 1497-1500.

Park, J.T., Li, M., Nakayama, K., Mao, T.L., Davidson, B., Zhang, Z., Kurman, R.J., Eberhart, C.G., Shih Ie, M., and Wang, T.L. 2006. Notch3 gene amplification in ovarian cancer. *Cancer Res* **66**(12): 6312-6318.

Pei, J., Balsara, B.R., Li, W., Litwin, S., Gabrielson, E., Feder, M., Jen, J., and Testa, J.R. 2001. Genomic imbalances in human lung adenocarcinomas and squamous cell carcinomas. *Genes Chromosomes Cancer* **31**(3): 282-287.

Petkov, P.M., Cassell, M.A., Sargent, E.E., Donnelly, C.J., Robinson, P., Crew, V., Asquith, S., Haar, R.V., and Wiles, M.V. 2004a. Development of a SNP genotyping panel for genetic monitoring of the laboratory mouse. *Genomics* **83**(5): 902-911.

Petkov, P.M., Ding, Y., Cassell, M.A., Zhang, W., Wagner, G., Sargent, E.E., Asquith, S., Crew, V., Johnson, K.A., Robinson, P., Scott, V.E., and Wiles, M.V. 2004b. An efficient SNP system for mouse genome scanning and elucidating strain relationships. *Genome Res* **14**(9): 1806-1811.

Pinkel, D. and Albertson, D.G. 2005. Array comparative genomic hybridization and its applications in cancer. *Nat Genet* **37 Suppl**: S11-17.

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. *Nat Genet* **20**(2): 207-211.

Pletcher, M.T., McClurg, P., Batalov, S., Su, A.I., Barnes, S.W., Lagler, E., Korstanje, R., Wang, X., Nusskern, D., Bogue, M.A., Mural, R.J., Paigen, B., and Wiltshire, T. 2004. Use of a dense single nucleotide polymorphism map for in silico mapping in the mouse. *PLoS Biol* **2**(12): e393.

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. *Nat Genet* **23**(1): 41-46.

Prince, J.A., Feuk, L., Howell, W.M., Jobs, M., Emahazion, T., Blennow, K., and Brookes, A.J. 2001. Robust and accurate single nucleotide polymorphism genotyping by dynamic allele-specific hybridization (DASH): design criteria and assay validation. *Genome Res* **11**(1): 152-162.

Pulciani, S., Santos, E., Long, L.K., Sorrentino, V., and Barbacid, M. 1985. ras gene Amplification and malignant transformation. *Mol Cell Biol* **5**(10): 2836-2841.

Radany, E.H., Hong, K., Kesharvarzi, S., Lander, E.S., and Bishop, J.M. 1997. Mouse mammary tumor virus/v-Ha-ras transgene-induced mammary tumors exhibit strain-specific allelic loss on mouse chromosome 4. *Proc Natl Acad Sci U S A* **94**(16): 8664-8669.

Ranade, K., Chang, M.S., Ting, C.T., Pei, D., Hsiao, C.F., Olivier, M., Pesich, R., Hebert, J., Chen, Y.D., Dzau, V.J., Curb, D., Olshen, R., Risch, N., Cox, D.R., and Botstein, D. 2001. High-throughput genotyping with single nucleotide polymorphisms. *Genome Res* **11**(7): 1262-1268.

Reilly, K.M., Tuskan, R.G., Christy, E., Loisel, D.A., Ledger, J., Bronson, R.T., Smith, C.D., Tsang, S., Munroe, D.J., and Jacks, T. 2004. Susceptibility to astrocytoma in mice mutant for Nf1 and Trp53 is linked to chromosome 11 and subject to epigenetic effects. *Proc Natl Acad Sci U S A* **101**(35): 13008-13013.

Ross, P., Hall, L., Smirnov, I., and Haff, L. 1998. High level multiplex genotyping by MALDI-TOF mass spectrometry. *Nat Biotechnol* **16**(13): 1347-1351.

Ross, P.L., Lee, K., and Belgrader, P. 1997. Discrimination of single-nucleotide polymorphisms in human DNA using peptide nucleic acid probes detected by MALDI-TOF mass spectrometry. *Anal Chem* **69**(20): 4197-4202.

Saiki, R.K., Chang, C.A., Levenson, C.H., Warren, T.C., Boehm, C.D., Kazazian, H.H., Jr., and Erlich, H.A. 1988. Diagnosis of sickle cell anemia and beta-thalassemia with enzymatically amplified DNA and nonradioactive allele-specific oligonucleotide probes. *N Engl J Med* **319**(9): 537-541.

Samuels, Y., Wang, Z., Bardelli, A., Silliman, N., Ptak, J., Szabo, S., Yan, H., Gazdar, A., Powell, S.M., Riggins, G.J., Willson, J.K., Markowitz, S., Kinzler, K.W., Vogelstein, B., and Velculescu, V.E. 2004. High frequency of mutations of the PIK3CA gene in human cancers. *Science* **304**(5670): 554.

Sasaki, M. 1982. Current status of cytogenetic studies in animal tumors with special reference to nonrandom chromosome changes. *Cancer Genet Cytogenet* **5**(2): 153-172.

Schaid, D.J., Guenther, J.C., Christensen, G.B., Hebbring, S., Rosenow, C., Hilker, C.A., McDonnell, S.K., Cunningham, J.M., Slager, S.L., Blute, M.L., and Thibodeau, S.N. 2004. Comparison of microsatellites versus single-nucleotide polymorphisms in a genome linkage screen for prostate cancer-susceptibility Loci. *Am J Hum Genet* **75**(6): 948-965.

Scholnick, S.B. and Richter, T.M. 2003. The role of CSMD1 in head and neck carcinogenesis. *Genes Chromosomes Cancer* **38**(3): 281-283.

Sellick, G.S., Webb, E.L., Allinson, R., Matutes, E., Dyer, M.J., Jonsson, V., Langerak, A.W., Mauro, F.R., Fuller, S., Wiley, J., Lyttelton, M., Callea, V., Yuille, M., Catovsky, D., and Houlston, R.S. 2005. A high-density SNP genomewide linkage scan for chronic lymphocytic leukemia-susceptibility loci. *Am J Hum Genet* **77**(3): 420-429.

Selzer, R.R., Richmond, T.A., Pofahl, N.J., Green, R.D., Eis, P.S., Nair, P., Brothman, A.R., and Stallings, R.L. 2005. Analysis of chromosome breakpoints in neuroblastoma at sub-kilobase resolution using fine-tiling oligonucleotide array CGH. *Genes Chromosomes Cancer* **44**(3): 305-319.

Shao, C., Stambrook, P.J., and Tischfield, J.A. 2001. Mitotic recombination is suppressed by chromosomal divergence in hybrids of distantly related mouse strains. *Nat Genet* **28**(2): 169-172.

Shen, R., Fan, J.B., Campbell, D., Chang, W., Chen, J., Doucet, D., Yeakley, J., Bibikova, M., Wickham Garcia, E., McBride, C., Steemers, F., Garcia, F., Kermani, B.G., Gunderson, K., and Oliphant, A. 2005. High-throughput SNP genotyping on universal bead arrays. *Mutat Res* **573**(1-2): 70-82.

Shih, C., Padhy, L.C., Murray, M., and Weinberg, R.A. 1981. Transforming genes of carcinomas and neuroblastomas introduced into mouse fibroblasts. *Nature* **290**(5803): 261-264.

Shih Ie, M., Sheu, J.J., Santillan, A., Nakayama, K., Yen, M.J., Bristow, R.E., Vang, R., Parmigiani, G., Kurman, R.J., Trope, C.G., Davidson, B., and Wang, T.L. 2005. Amplification of a chromatin remodeling gene, Rsf-1/HBXAP, in ovarian carcinoma. *Proc Natl Acad Sci U S A* **102**(39): 14004-14009.

Shimkin, M.B. 1977. *Contrary to nature: being an illustrated commentary on some persons and events of historical importance in the development of knowledge concerning. cancer*. U.S. Dept. of Health, Education, and Welfare, Public Health Service, National Institutes of Health: for sale by Supt. of Docs., U.S. Govt. Print. Off., [Washington].

Shumaker, J.M., Metspalu, A., and Caskey, C.T. 1996. Mutation detection by solid phase primer extension. *Hum Mutat* **7**(4): 346-354.

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. *Nat Genet* **29**(3): 263-264.

Snijders, A.M., Nowak, N.J., Huey, B., Fridlyand, J., Law, S., Conroy, J., Tokuyasu, T., Demir, K., Chiu, R., Mao, J.H., Jain, A.N., Jones, S.J., Balmain, A., Pinkel, D., and Albertson, D.G. 2005. Mapping segmental and sequence variations among laboratory mice using BAC array CGH. *Genome Res* **15**(2): 302-311.

Sokolov, B.P. 1990. Primer extension technique for the detection of single nucleotide in genomic DNA. *Nucleic Acids Res* **18**(12): 3671.

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**(4): 399-407.

Soriano, P. 1999. Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nat Genet* **21**(1): 70-71.

Sorrentino, V., McKinney, M.D., Drozdoff, V., Hume, C.R., and Fleissner, E. 1988. Spontaneous or carcinogen-mediated amplification of a mutated ras gene promotes neoplastic transformation. *Oncogene Res* **2**(2): 189-195.

Sosnowski, R.G., Tu, E., Butler, W.F., O'Connell, J.P., and Heller, M.J. 1997. Rapid determination of single base mismatch mutations in DNA hybrids by direct electric field control. *Proc Natl Acad Sci U S A* **94**(4): 1119-1123.

Southern, E.M., Maskos, U., and Elder, J.K. 1992. Analyzing and comparing nucleic acid sequences by hybridization to arrays of oligonucleotides: evaluation using experimental models. *Genomics* **13**(4): 1008-1017.

Stephens, P., Hunter, C., Bignell, G., Edkins, S., Davies, H., Teague, J., Stevens, C., O'Meara, S., Smith, R., Parker, A., Barthorpe, A., Blow, M., Brackenbury, L., Butler, A., Clarke, O., Cole, J., Dicks, E., Dike, A., Drozd, A., Edwards, K., Forbes, S., Foster, R., Gray, K., Greenman, C., Halliday, K., Hills, K., Kosmidou, V., Lugg, R., Menzies, A., Perry, J., Petty, R., Raine, K., Ratford, L., Shepherd, R., Small, A., Stephens, Y., Tofts, C., Varian, J., West, S., Widaa, S., Yates, A., Brasseur, F., Cooper, C.S., Flanagan, A.M., Knowles, M., Leung, S.Y., Louis,

D.N., Looijenga, L.H., Malkowicz, B., Pierotti, M.A., Teh, B., Chenevix-Trench, G., Weber, B.L., Yuen, S.T., Harris, G., Goldstraw, P., Nicholson, A.G., Futreal, P.A., Wooster, R., and Stratton, M.R. 2004. Lung cancer: intragenic ERBB2 kinase mutations in tumours. *Nature* **431**(7008): 525-526.

Sweet-Cordero, A., Mukherjee, S., Subramanian, A., You, H., Roix, J.J., Ladd-Acosta, C., Mesirov, J., Golub, T.R., and Jacks, T. 2005. An oncogenic KRAS2 expression signature identified by cross-species gene-expression analysis. *Nat Genet* **37**(1): 48-55.

Sweet-Cordero, A., Tseng, G.C., You, H., Douglass, M., Huey, B., Albertson, D., and Jacks, T. 2006. Comparison of gene expression and DNA copy number changes in a murine model of lung cancer. *Genes Chromosomes Cancer* **45**(4): 338-348.

Syvanen, A.C. 1999. From gels to chips: "minisequencing" primer extension for analysis of point mutations and single nucleotide polymorphisms. *Hum Mutat* **13**(1): 1-10.

-. 2001. Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nat Rev Genet* **2**(12): 930-942.

Syvanen, A.C., Aalto-Setala, K., Harju, L., Kontula, K., and Soderlund, H. 1990. A primer-guided nucleotide incorporation assay in the genotyping of apolipoprotein E. *Genomics* **8**(4): 684-692.

Tai, A.L., Mak, W., Ng, P.K., Chua, D.T., Ng, M.Y., Fu, L., Chu, K.K., Fang, Y., Qiang Song, Y., Chen, M., Zhang, M., Sham, P.C., and Guan, X.Y. 2006. High-throughput loss-of-heterozygosity study of chromosome 3p in lung cancer using single-nucleotide polymorphism markers. *Cancer Res* **66**(8): 4133-4138.

Taton, T.A., Mirkin, C.A., and Letsinger, R.L. 2000. Scanometric DNA array detection with nanoparticle probes. *Science* **289**(5485): 1757-1760.

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**(3): 718-725.

Testa, J.R., Siegfried, J.M., Liu, Z., Hunt, J.D., Feder, M.M., Litwin, S., Zhou, J.Y., Taguchi, T., and Keller, S.M. 1994. Cytogenetic analysis of 63 non-small cell lung carcinomas: recurrent chromosome alterations amid frequent and widespread genomic upheaval. *Genes Chromosomes Cancer* **11**(3): 178-194.

Thomas, R., Scott, A., Langford, C.F., Fosmire, S.P., Jubala, C.M., Lorentzen, T.D., Hitte, C., Karlsson, E.K., Kirkness, E., Ostrander, E.A., Galibert, F., Lindblad-Toh, K., Modiano, J.F., and Breen, M. 2005. Construction of a 2-Mb resolution BAC microarray for CGH analysis of canine tumors. *Genome Res* **15**(12): 1831-1837.

Thomas, R., Smith, K.C., Ostrander, E.A., Galibert, F., and Breen, M. 2003. Chromosome aberrations in canine multicentric lymphomas detected with comparative genomic hybridisation and a panel of single locus probes. *Br J Cancer* **89**(8): 1530-1537.

Tonon, G., Wong, K.K., Maulik, G., Brennan, C., Feng, B., Zhang, Y., Khatry, D.B., Protopopov, A., You, M.J., Aguirre, A.J., Martin, E.S., Yang, Z., Ji, H., Chin, L., and Depinho, R.A. 2005. High-resolution genomic profiles of human lung cancer. *Proc Natl Acad Sci U S A* **102**(27): 9625-9630.

Trkova, M., Foretova, L., Kodet, R., Hedvicakova, P., and Sedlacek, Z. 2003. A Li-Fraumeni syndrome family with retained heterozygosity for a germline TP53 mutation in two tumors. *Cancer Genet Cytogenet* **145**(1): 60-64.

Tyagi, S., Bratu, D.P., and Kramer, F.R. 1998. Multicolor molecular beacons for allele discrimination. *Nat Biotechnol* **16**(1): 49-53.

Tyagi, S. and Kramer, F.R. 1996. Molecular beacons: probes that fluoresce upon hybridization. *Nat Biotechnol* **14**(3): 303-308.

van den Ijssel, P., Tijssen, M., Chin, S.F., Eijk, P., Carvalho, B., Hopmans, E., Holstege, H., Bangarusamy, D.K., Jonkers, J., Meijer, G.A., Caldas, C., and Ylstra, B. 2005. Human and mouse oligonucleotide-based array CGH. *Nucleic Acids Res* **33**(22): e192.

Van Dyke, T. and Jacks, T. 2002. Cancer modeling in the modern era: progress and challenges. *Cell* **108**(2): 135-144.

Velculescu, V.E., Zhang, L., Vogelstein, B., and Kinzler, K.W. 1995. Serial analysis of gene expression. *Science* **270**(5235): 484-487.

Venkatachalam, S., Shi, Y.P., Jones, S.N., Vogel, H., Bradley, A., Pinkel, D., and Donehower, L.A. 1998. Retention of wild-type p53 in tumors from p53

heterozygous mice: reduction of p53 dosage can promote cancer formation. *Embo J* **17**(16): 4657-4667.

von Ahsen, N., Oellerich, M., Armstrong, V.W., and Schutz, E. 1999. Application of a thermodynamic nearest-neighbor model to estimate nucleic acid stability and optimize probe design: prediction of melting points of multiple mutations of apolipoprotein B-3500 and factor V with a hybridization probe genotyping assay on the LightCycler. *Clin Chem* **45**(12): 2094-2101.

Wade, C.M., Kulbokas, E.J., 3rd, 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**(6915): 574-578.

Waggoner, A. 2006. Fluorescent labels for proteomics and genomics. *Curr Opin Chem Biol* **10**(1): 62-66.

Wang, T.L., Maierhofer, C., Speicher, M.R., Lengauer, C., Vogelstein, B., Kinzler, K.W., and Velculescu, V.E. 2002. Digital karyotyping. *Proc Natl Acad Sci U S A* **99**(25): 16156-16161.

Wang, Z., Shen, D., Parsons, D.W., Bardelli, A., Sager, J., Szabo, S., Ptak, J., Silliman, N., Peters, B.A., van der Heijden, M.S., Parmigiani, G., Yan, H., Wang, T.L., Riggins, G., Powell, S.M., Willson, J.K., Markowitz, S., Kinzler, K.W., Vogelstein, B., and Velculescu, V.E. 2004. Mutational analysis of the tyrosine phosphatome in colorectal cancers. *Science* **304**(5674): 1164-1166.

Waterfall, C.M. and Cobb, B.D. 2001. Single tube genotyping of sickle cell anaemia using PCR-based SNP analysis. *Nucleic Acids Res* **29**(23): E119.

Waterston, R.H. Lindblad-Toh, K. Birney, E. Rogers, J. Abril, J.F. Agarwal, P. Agarwala, R. Ainscough, R. Alexandersson, M. An, P. Antonarakis, S.E. Attwood, J. Baertsch, R. Bailey, J. Barlow, K. Beck, S. Berry, E. Birren, B. Bloom, T. Bork, P. Botcherby, M. Bray, N. Brent, M.R. Brown, D.G. Brown, S.D. Bult, C. Burton, J. Butler, J. Campbell, R.D. Carninci, P. Cawley, S. Chiaromonte, F. Chinwalla, A.T. Church, D.M. Clamp, M. Clee, C. Collins, F.S. Cook, L.L. Copley, R.R. Coulson, A. Couronne, O. Cuff, J. Curwen, V. Cutts, T. Daly, M. David, R. Davies, J. Delehaunty, K.D. Deri, J. Dermitzakis, E.T. Dewey, C. Dickens, N.J. Diekhans, M. Dodge, S. Dubchak, I. Dunn, D.M. Eddy, S.R. Elnitski, L. Emes, R.D. Eswara, P. Eyras, E. Felsenfeld, A. Fewell, G.A. Flicek, P. Foley, K. Frankel, W.N. Fulton, L.A. Fulton, R.S. Furey, T.S. Gage, D. Gibbs, R.A. Glusman, G. Gnerre, S. Goldman, N. Goodstadt, L. Grafham, D.

Graves, T.A. Green, E.D. Gregory, S. Guigo, R. Guyer, M. Hardison, R.C. Haussler, D. Hayashizaki, Y. Hillier, L.W. Hinrichs, A. Hlavina, W. Holzer, T. Hsu, F. Hua, A. Hubbard, T. Hunt, A. Jackson, I. Jaffe, D.B. Johnson, L.S. Jones, M. Jones, T.A. Joy, A. Kamal, M. Karlsson, E.K. Karolchik, D. Kasprzyk, A. Kawai, J. Keibler, E. Kells, C. Kent, W.J. Kirby, A. Kolbe, D.L. Korf, I. Kucherlapati, R.S. Kulbokas, E.J. Kulp, D. Landers, T. Leger, J.P. Leonard, S. Letunic, I. Levine, R. Li, J. Li, M. Lloyd, C. Lucas, S. Ma, B. Maglott, D.R. Mardis, E.R. Matthews, L. Mauceli, E. Mayer, J.H. McCarthy, M. McCombie, W.R. McLaren, S. McLay, K. McPherson, J.D. Meldrim, J. Meredith, B. Mesirov, J.P. Miller, W. Miner, T.L. Mongin, E. Montgomery, K.T. Morgan, M. Mott, R. Mullikin, J.C. Muzny, D.M. Nash, W.E. Nelson, J.O. Nhan, M.N. Nicol, R. Ning, Z. Nusbaum, C. O'Connor, M.J. Okazaki, Y. Oliver, K. Overton-Larty, E. Pachter, L. Parra, G. Pepin, K.H. Peterson, J. Pevzner, P. Plumb, R. Pohl, C.S. Poliakov, A. Ponce, T.C. Ponting, C.P. Potter, S. Quail, M. Reymond, A. Roe, B.A. Roskin, K.M. Rubin, E.M. Rust, A.G. Santos, R. Sapojnikov, V. Schultz, B. Schultz, J. Schwartz, M.S. Schwartz, S. Scott, C. Seaman, S. Searle, S. Sharpe, T. Sheridan, A. Shownkeen, R. Sims, S. Singer, J.B. Slater, G. Smit, A. Smith, D.R. Spencer, B. Stabenau, A. Stange-Thomann, N. Sugnet, C. Suyama, M. Tesler, G. Thompson, J. Torrents, D. Trevaskis, E. Tromp, J. Ucla, C. Ureta-Vidal, A. Vinson, J.P. Von Niederhausern, A.C. Wade, C.M. Wall, M. Weber, R.J. Weiss, R.B. Wendl, M.C. West, A.P. Wetterstrand, K. Wheeler, R. Whelan, S. Wierzbowski, J. Willey, D. Williams, S. Wilson, R.K. Winter, E. Worley, K.C. Wyman, D. Yang, S. Yang, S.P. Zdobnov, E.M. Zody, M.C. and Lander, E.S. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**(6915): 520-562.

Weiss, W.A., Aldape, K., Mohapatra, G., Feuerstein, B.G., and Bishop, J.M. 1997. Targeted expression of MYCN causes neuroblastoma in transgenic mice. *Embo J* **16**(11): 2985-2995.

Wilhelm, M., Veltman, J.A., Olshen, A.B., Jain, A.N., Moore, D.H., Presti, J.C., Jr., Kovacs, G., and Waldman, F.M. 2002. Array-based comparative genomic hybridization for the differential diagnosis of renal cell cancer. *Cancer Res* **62**(4): 957-960.

Wiltshire, T., Pletcher, M.T., Batalov, S., Barnes, S.W., Tarantino, L.M., Cooke, M.P., Wu, H., Smylie, K., Santrosyan, A., Copeland, N.G., Jenkins, N.A., Kalush, F., Mural, R.J., Glynne, R.J., Kay, S.A., Adams, M.D., and Fletcher, C.F. 2003. Genome-wide single-nucleotide polymorphism analysis defines haplotype patterns in mouse. *Proc Natl Acad Sci U S A* **100**(6): 3380-3385.

Wong, M.P., Fung, L.F., Wang, E., Chow, W.S., Chiu, S.W., Lam, W.K., Ho, K.K., Ma, E.S., Wan, T.S., and Chung, L.P. 2003. Chromosomal aberrations of primary lung adenocarcinomas in nonsmokers. *Cancer* **97**(5): 1263-1270.

Wu, Y., Renard, C.A., Apiou, F., Huerre, M., Tiollais, P., Dutrillaux, B., and Buendia, M.A. 2002. Recurrent allelic deletions at mouse chromosomes 4 and 14 in Myc-induced liver tumors. *Oncogene* **21**(10): 1518-1526.

Xu, H., Sha, M.Y., Wong, E.Y., Uphoff, J., Xu, Y., Treadway, J.A., Truong, A., O'Brien, E., Asquith, S., Stubbins, M., Spurr, N.K., Lai, E.H., and Mahoney, W. 2003. Multiplexed SNP genotyping using the Qbead system: a quantum dot-encoded microsphere-based assay. *Nucleic Acids Res* **31**(8): e43.

Ylstra, B., van den Ijssel, P., Carvalho, B., Brakenhoff, R.H., and Meijer, G.A. 2006. BAC to the future! or oligonucleotides: a perspective for micro array comparative genomic hybridization (array CGH). *Nucleic Acids Res* **34**(2): 445-450.

You, M.J., Castrillon, D.H., Bastian, B.C., O'Hagan, R.C., Bosenberg, M.W., Parsons, R., Chin, L., and DePinho, R.A. 2002. Genetic analysis of Pten and Ink4a/Arf interactions in the suppression of tumorigenesis in mice. *Proc Natl Acad Sci U S A* **99**(3): 1455-1460.

Zabarovsky, E.R., Lerman, M.I., and Minna, J.D. 2002. Tumor suppressor genes on chromosome 3p involved in the pathogenesis of lung and other cancers. *Oncogene* **21**(45): 6915-6935.

Zhang, J., Hunter, K.W., Gandolph, M., Rowe, W.L., Finney, R.P., Kelley, J.M., Edmonson, M., and Buetow, K.H. 2005. A high-resolution multistrain haplotype analysis of laboratory mouse genome reveals three distinctive genetic variation patterns. *Genome Res* **15**(2): 241-249.

Zhang, Z., Wang, Y., Vikis, H.G., Johnson, L., Liu, G., Li, J., Anderson, M.W., Sills, R.C., Hong, H.L., Devereux, T.R., Jacks, T., Guan, K.L., and You, M. 2001. Wildtype Kras2 can inhibit lung carcinogenesis in mice. *Nat Genet* **29**(1): 25-33.

Zhao, X., Li, C., Paez, J.G., Chin, K., Janne, P.A., Chen, T.H., Girard, L., Minna, J., Christiani, D., Leo, C., Gray, J.W., Sellers, W.R., and Meyerson, M. 2004. An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays. *Cancer Res* **64**(9): 3060-3071.

Zhao, X., Weir, B.A., LaFramboise, T., Lin, M., Beroukhim, R., Garraway, L., Beheshti, J., Lee, J.C., Naoki, K., Richards, W.G., Sugarbaker, D., Chen, F., Rubin, M.A., Janne, P.A., Girard, L., Minna, J., Christiani, D., Li, C., Sellers, W.R., and Meyerson, M. 2005. Homozygous deletions and chromosome amplifications in

human lung carcinomas revealed by single nucleotide polymorphism array analysis. *Cancer Res* **65**(13): 5561-5570.

Zielinski, B., Gratias, S., Toedt, G., Mendrzyk, F., Stange, D.E., Radlwimmer, B., Lohmann, D.R., and Lichter, P. 2005. Detection of chromosomal imbalances in retinoblastoma by matrix-based comparative genomic hybridization. *Genes Chromosomes Cancer* **43**(3): 294-301.

Chapter 2

Mouse Representational Oligonucleotide Microarray Analysis Detects Copy Number Alterations in Murine Tumors of the Lung and Retina

Mandy Tam[1], David MacPherson[1], Alice Shaw[1], David Mu[2], and Tyler Jacks[1,3]

[1]Massachusetts Institute of Technology, Department of Biology and Center for Cancer Research, 77 Massachusetts Avenue, Cambridge, MA 02139 USA

[2]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA

[3]Howard Hughes Medical Institute, 4000 James Bridge Road, Chevy Chase, MD 20185, USA

The author prepared lung tumor DNA samples for analysis and performed Southern blotting validation of the retinoblastoma data. Retinoblastoma DNA was prepared by D. Macpherson. ROMA hybridizations were performed by D.Mu. The Kras mouse colony was maintained by A. Shaw.

Abstract

A wide range of copy number alterations in primary mouse tumors has been previously documented, including single copy gain or loss of entire chromosomes, partial gain or loss of a chromosome, high-amplitude focal amplifications, to low-level small deletions. To cover this broad spectrum, a genome-wide high-resolution CGH tool would be invaluable for mouse cancer DNA anlayses. Representational Oligonucleotide Microarray Analysis (ROMA) employs an integrated genome complexity reduction step that can enhance signal to noise ratio during genome hybridization. Among the many CGH platforms for genome-wide copy number analysis studies in human, ROMA has one of the highest resolving power averaging at 30kb (Lucito et al. 2003). We tested the mouse version of ROMA on mouse retinoblasomas and lung adenomas. We were able to detect a focal high-amplitude (>4.6fold) *N-Myc* amplification in retinoblastomas of a Rb/p130 DKO model, as well as numerous whole-chromosomal gains and losses in the same retinoblastoma sample set and in the lung tumors driven by a *Kras* mutation.

Introduction

Comparative Genomic Hybridization (CGH) is a technique that measures changes in the amount of DNA throughout the whole genome. Chromosomal copy number imbalances are commonly observed in cancer, particularly in carcinomas and these imbalances are detectable by CGH. The format of CGH has evolved from using metaphase spreads to microarrays as hybridization for differentially labeled tumor and normal genomes (Kallioniemi et al. 1992). Several microarray formats have been developed using probes that range from large-insert clones such as BACs (bacterial artificial chromosomes), PACs (P1 artificial chromosomes), or YACs (yeast artificial chromosomes), single-stranded DNA, cDNAs, and oligonucleotides (Solinas-Toldo et al. 1997; Pinkel et al. 1998; Pollack et al. 1999; Barrett et al. 2004; Dhami et al. 2005). Over the past few years, BAC-based and oligonucleotide-based arrays predominate the race for better resolution. Representational Oligonucleotide MicroArray (ROMA) is among one of the competing technologies that has proven useful in detecting genetic lesions in cancer (Lucito et al. 2003). ROMA involves the generation of low-complexity representations of the genomes to reduce hybridization noise. A sub-Mb resolution can be achieved (Lucito et al. 2003).

In many human cancers, large-scale chromosomal abnormalities are common but identification of the critical regions is often difficult. Comparative genomic studies between mouse and human may be one way to aid this search. Individual chromosomes in the mouse can be delimited into regions of synteny that are conserved in human on separate chromosomes. Given this structure and the high overall conservation between

mouse and human, genomic studies in mice may provide interesting insights into cancer-associated genetic lesions in human.

The high-resolution ROMA technology has been translated for use in the mouse. In the present study, we utilized this platform to characterize two different types of mouse cancer: 1) retinoblastomas in mice doubly deleted for *Rb* and *p130* in the retina (MacPherson et al. submitted), and 2) non-small cell lung cancer (NSCLC) in mice conditionally expressing a $Kras^{G12D}$ mutant allele in the lung (Jackson et al. 2005). We sought to characterize the genetic alterations that may cooperate with these initiating genetic lesions. In humans, recurrent large-scale chromosomal imbalances have been observed. In retinoblastomas, chromosomal gains include those in 6p, 1q, 2p and loss is frequently seen in 16q (Mairal et al. 2000; Chen et al. 2001; Lillington et al. 2003; Zielinski et al. 2005). In NSCLC, chromosomal gains in 1q, 3q, 5p, 8q, and losses in 3p, 8p, 9q, 13q, 17p have been identified (Balsara and Testa 2002). In this chapter, we report the usefulness of ROMA in identifying single gene amplification as well as whole chromosomal changes in tumor-derived DNA from these mouse models.

Results

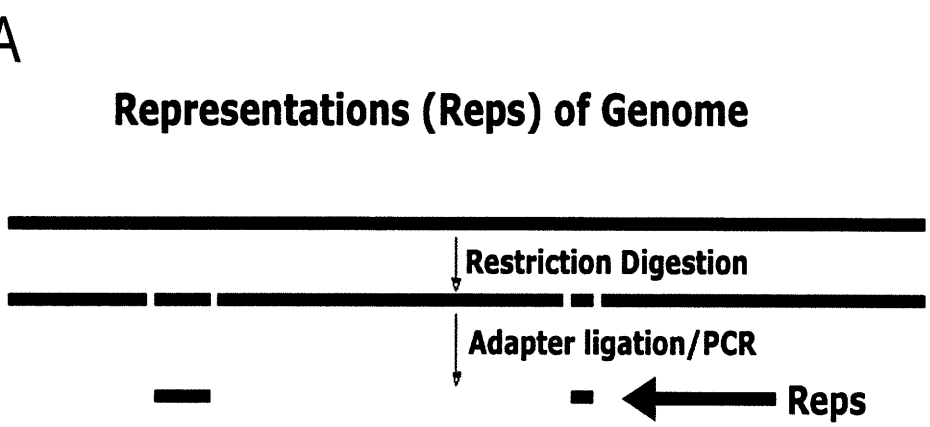
Representational oligonucleotide microarray analysis (ROMA) is a tool to detect copy number changes in the genome (Lucito et al. 2003). The technique involves the hybridization of low-complexity representations of tumor vs. normal genomes to oligonucleotide probes on microarrays (Figure 1). The usefulness of the ROMA platform to detect DNA copy number changes has been demonstrated (Jobanputra et al. 2005).

Figure 1: Schematic for Representational Oligonucleotide MicroArray (ROMA) analysis

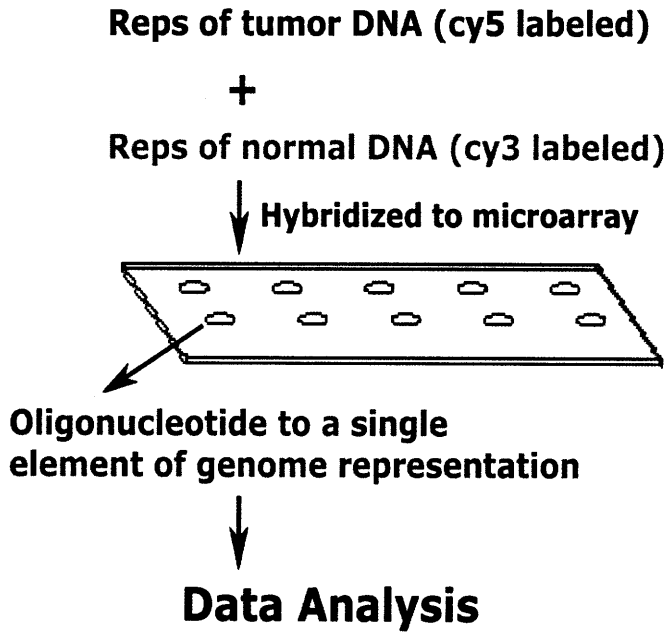
A) Generation of low complexity representations (LCR) of the genomes. Tumor and normal DNA was digested with BglII and amplified by linker-based PCR.

B) Tumor and normal LCRs were differentially labeled with Cy-5 and Cy-3 fluorescent dyes. Hybridization of the tumor vs. normal LCRs was performed on 84K oligonucleotide arrays. Each oligonucleotide was designed to bind selected *Bgl*II fragments in the mouse genome.

A


Representations (Reps) of Genome
Restriction Digestion
Adapter ligation/PCR
Reps

B


Reps of tumor DNA (cy5 labeled)
+
Reps of normal DNA (cy3 labeled)
Hybridized to microarray
Oligonucleotide to a single
element of genome representation
Data Analysis

Groups in Cold Spring Harbor Laboratory and Nimblegen Systems Inc. have since translated the ROMA technique for genomic analysis in the mouse. This study reports the implementation of ROMA to characterize copy number changes in tumor DNA from two different mouse cancer models.

Detection of N-Myc amplification in mouse retinoblastomas

Retinoblastomas can be induced in mice with a retina-specific Rb deletion combined with an inactivation of *p130* (MacPherson et al., submitted). As described by the authors, in this *Rb/p130* DKO model, tumors could be consistently observed with a latency of 128±18 days (means±s.d.). Early tumors arose in the periphery of the retina by PND21 when retinal development was completed. The tumors continued to progress in the adult mice and filled the posterior and anterior chambers of the eyes. At the experimental end point, tumors cells could be seen infiltrating the optic nerve. In addition, metastases to lymph nodes were observed in 38% of the animals.

To identify genetic lesions that cooperate with *Rb* and *p130* deletions in this retinoblastoma model, ROMA was performed on 8 lymph node metastases. Observed regions of copy number gain and loss are summarized in Table 1. Recurrent changes included whole chromosome 1 and chromosome 12 gains, which was each found in 4/8 tumors. In addition, focal amplifications in 12qA1.1 was detected in DNA from three tumors: 9806, 4836 and Drb13. The amplicon sizes in the respective order were 1.9MB, 3.3Mb and 451kb with a 136kb minimal overlap, which harbors the N-myc oncogene (Figure 2A and B). Amplification of *N-Myc* was verified by Southern blotting. 3/6 tumor DNA samples in the Southern analysis have been profiled by ROMA, including those of

Table 1. Summary of chromosomal changes in 9 metastatic retinoblastomas from Rb/p130DKO mice.

tumor ID	gain	amplification	loss
9806	1, 12	**12qA1.1**, 12qF2	
7217	12		
4834	10qA4qter		2, 12, 18, 9qA5.3qter, 4qB3qter
4726	1		
4848	1, 12, 19		
4827		12qF2, 12qC1	
4836a*	1	**12qA1.1**, 3qf3, 12qF1-2	3qA3, 17qe2, 17qe1.1
drb13*	12	**12qA1.1**	

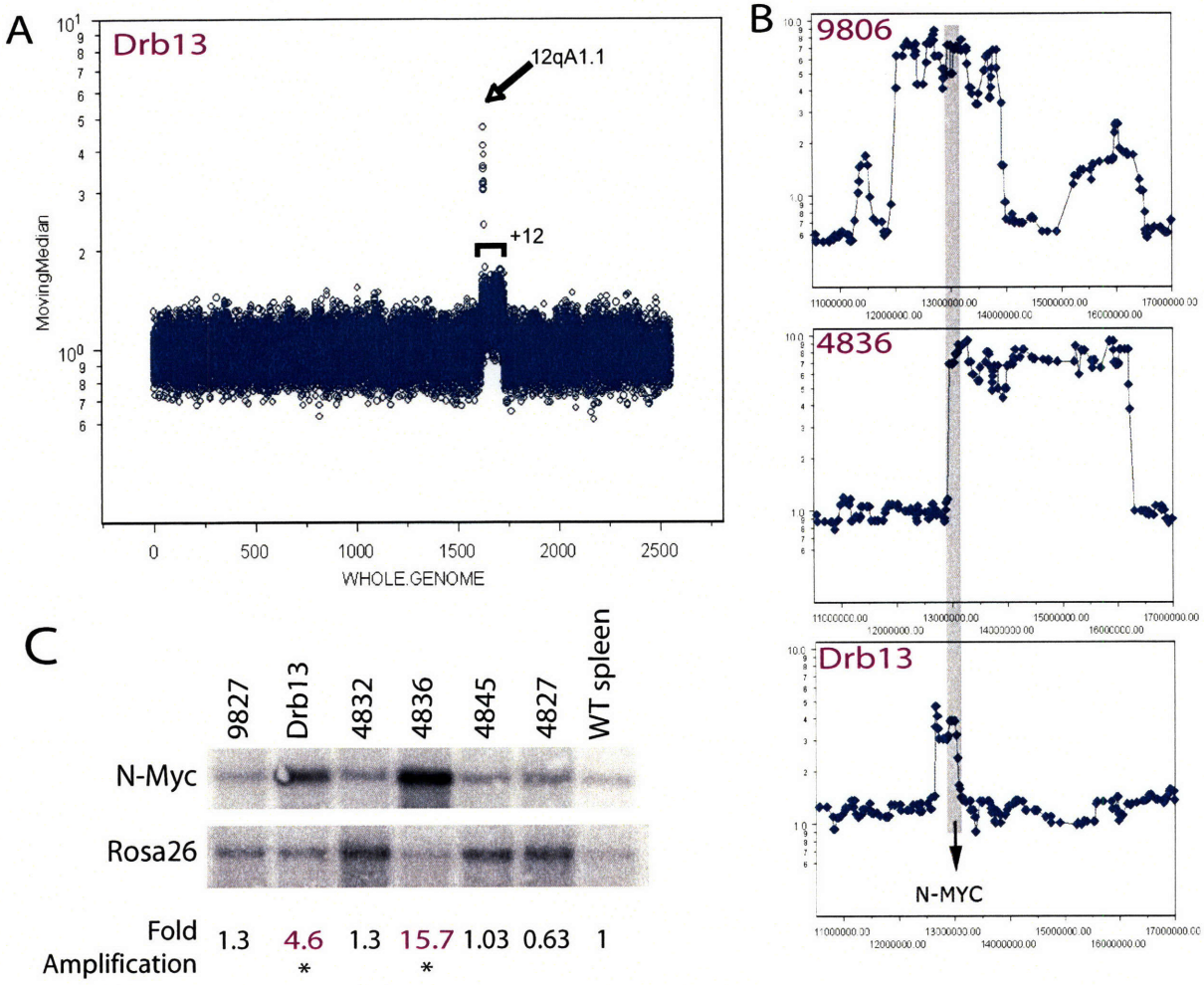
Amplicons at 12qA1.1 harboring N-myc gene are in bold
Samples selected for ROMA analysis based on presence of N-myc amplification detected by Southern blot
*Tail DNA used for ROMA hybridization was not from the tumor-containing mouse, thus, polymorphisms could contribute to focal changes

Figure 2: Detection of N-Myc amplification in metastatic retinoblastomas

A) Whole-genome copy number ratio plot of Drb13 tumor vs. normal DNA. The Y-axis depicts the moving median fluorescence ratios of Cy5 labeled tumor to Cy3 labeled tail DNA. The X-axis is an index of the probes genomic order based on the UCSC mouse May 2004 (mm5) annotated assembly. Whole chromosome 12 is increased in copy ratio (labeled). Within the chromosome, a focally amplified region is also observed (arrow) and is mapped to 12qA1.1.

B) A minimally overlapping region of amplification at 12qA1.1 in 3 retinoblastomas (9806, 4836, and drb13). The 136kb core amplicon is highlighted in the zoomed-in copy ratio plots of the 3 samples. The X-axis marks the nucleotide coordinates on chromosome 12 and the Y-axis is fluorescence ratio. *N-Myc* is the only known RefSeq gene residing in the area.

C) Southern analysis of *N-Myc* locus on DNA from 6 tumors. 3/6 tumors (Drb13, 4836, and 4827) have been profiled with ROMA. To control for sample loading, blot was re-probed for *Rosa26* locus on chromosome 6, a chromosome that did not show any copy number changes in the retinoblastoma samples. Relative signal of *N-Myc* to *Rosa26* probes was quantified on phosphorimages with the ImageQuant software. Using wild-type spleen DNA as normal, fold of *Myc* amplification is shown. Samples with high-magnitude amplification are marked by asterisks.


A
Drb13
12qA1.1
+12
MovingMedian
WHOLE.GENOME
B
9806
4836
Drb13
N-MYC
C
9827
Drb13
4832
4836
4845
4827
WT spleen
N-Myc
Rosa26
Fold Amplification
1.3
4.6
1.3
15.7
1.03
0.63
1
*
*

Drb13 and 4836. A 4.6 and 15.7-fold increase of N-Myc copy number was seen in Drb13 and 4836 tumor DNA respectively, confirming amplification data from ROMA (Figure 2C).

Detection of recurrent whole chromosomal changes in primary lung tumors

We performed a CGH study to identify genetic lesions that collaborate with a *Kras* mutation during lung cancer development in the mouse. We also assessed if a germline *p53* mutation or loss can lead to different mutation spectra in these tumors.

An inducible $Kras^{G12D}$ mouse model was used for the study. To induce lung tumors, $Kras^{LSLG12D}$;*p53*+/+, $Kras^{LSLG12D}$;*p53*+/- and $Kras^{LSLG12D}$;$p53^{R270H}$/+ mice were intranasally infected with adenovirus Cre. 28 primary lung tumors were obtained from animals between 22-24 weeks of age. The majority of the tumors were early-stage with uniform nuclei, graded 1-2 according to the scale described in Jackson et al. 2005. As shown in Table 2, only 3/28 tumors contained grade 3 characteristics and 1/28 tumor was graded 4. The histological distribution is comparable between tumors coming from mice of the 3 genotypes.

ROMA was performed to assess copy number alterations in tumors from mice of the 3 different genotypes. Tumor and normal DNA pairs were subjected to ROMA. Copy number gains and/or losses were observed in 10 out of 28 tumors. These changes graded 2 or higher. Among the changes as summarized in Figure 3, whole chromosomal copy alterations composed the majority. Chromosome 6 gain was the most frequent, found in 8/10 tumors that contained any changes. The second most common copy number

Table 2: Characteristics of lung tumors analyzed

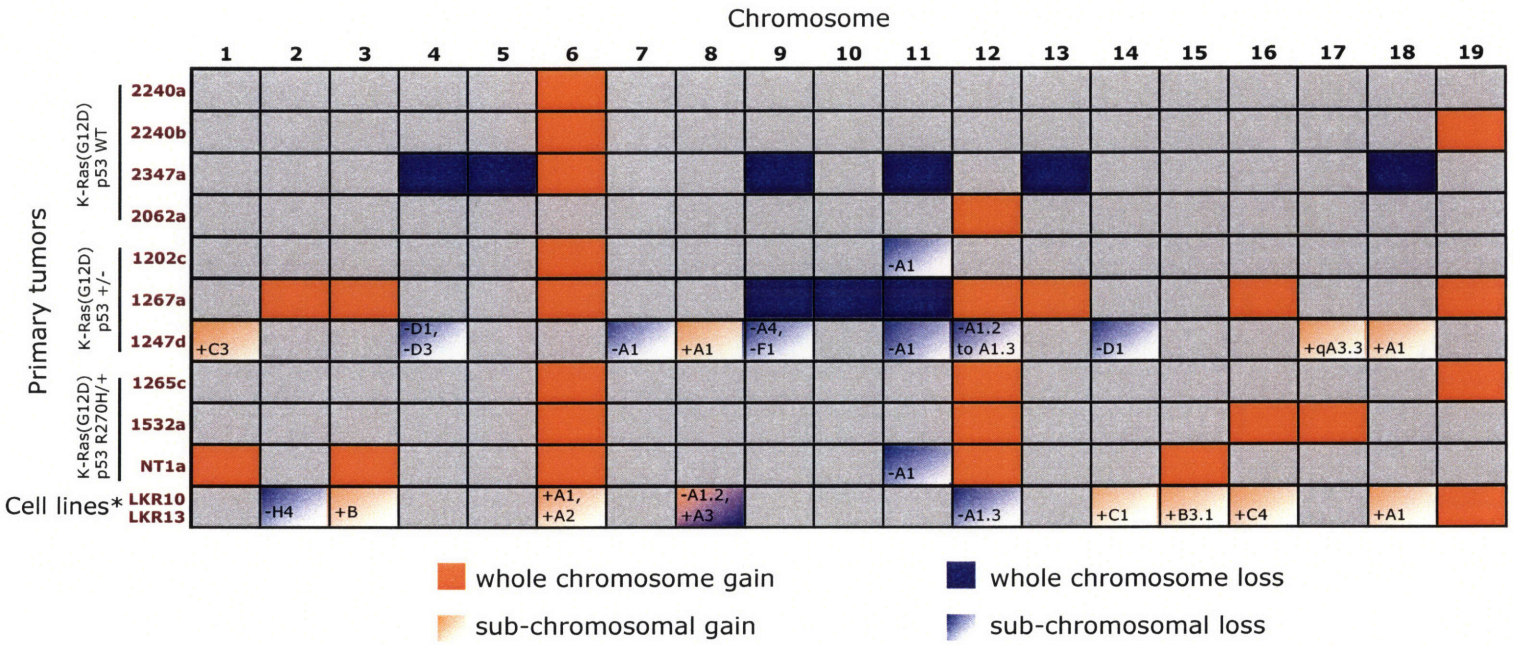
The table lists the genotype of the mice from which individual lung tumors were dissected out. Histological grading of the tumors was assigned using the criteria described previously (Jackson et al. 2005). Tumors that exhibited any copy number changes by ROMA are highlighted.

Tumor ID	Mouse genotype	Histology
1186a1	$Kras^{G12D}$	Grade 0-1, some normal tissue attached
1186b1	$Kras^{G12D}$	Grade 0-1, some normal tissue attached
1302b	$Kras^{G12D}$	Grade 1
1327a	$Kras^{G12D}$	Grade 1
1327b	$Kras^{G12D}$	Grade 1
1249b	$Kras^{G12D}$	Grade 2
2240a	$Kras^{G12D}$	Grade 2
2347a	$Kras^{G12D}$	Grade 2
2062a	$Kras^{G12D}$	Grade 2+
2240b	$Kras^{G12D}$	Grade 2+
1278b	$Kras^{G12D}$, $p53^{+/-}$	Grade 1
1275a	$Kras^{G12D}$, $p53^{+/-}$	Grade 1-2
1248a	$Kras^{G12D}$, $p53^{+/-}$	Grade 2
1278a	$Kras^{G12D}$, $p53^{+/-}$	Grade 2
1247d	$Kras^{G12D}$, $p53^{+/-}$	Grade 2-2+
1267a	$Kras^{G12D}$, $p53^{+/-}$	Grade 3, papillary
1202c	$Kras^{G12D}$, $p53^{+/-}$	Grade 4
1232b	$Kras^{G12D}$, $p53^{R270H/+}$	Grade 1
1265a	$Kras^{G12D}$, $p53^{R270H/+}$	Grade 1-2
1232c	$Kras^{G12D}$, $p53^{R270H/+}$	Grade 1-2
NT1a	$Kras^{G12D}$, $p53^{R270H/+}$	Grade 1-2
1232a	$Kras^{G12D}$, $p53^{R270H/+}$	Grade 2
1291b	$Kras^{G12D}$, $p53^{R270H/+}$	Grade 2, infiltrating lymphocytes
1265c	$Kras^{G12D}$, $p53^{R270H/+}$	Grade 2
1291a	$Kras^{G12D}$, $p53^{R270H/+}$	Grade 2+
1782a	$Kras^{G12D}$, $p53^{R270H/+}$	Grade 2+-3
1532a	$Kras^{G12D}$, $p53^{R270H/+}$	Grade 2-3
1782b	$Kras^{G12D}$, $p53^{R270H/+}$	n/a
LKR10	$Kras^{LA1}$*	tumor derived cell line
LKR13	$Kras^{LA1}$*	tumor derived cell line

*LKR10 and LKR13 cell lines were derived from the same KrasLA1 mouse.

Figure 3: Summary of genetic changes seen primary lung tumors and lung tumor-derived cell lines

Genetic changes observed in 9 of the tested tumors and 2 cell lines are summarized in a grid. Each row is an individual sample. Genotypes of mice that gave rise to the tumors were indicated. The two independently maintained cell lines LKR10 and LKR13 from the same mouse had almost identical ROMA profiles. The cell lines likely have the same clonal origin, thus their data are presented in same row (*) of this grid. The columns of the grid represent different chromosomes. Grey boxes indicate silence; solid orange means whole chromosomal gain; solid blue is whole chromosomal loss. Sub-chromosomal gains and losses are illustrated with shaded orange and blue respectively, and the region(s) altered are labeled inside the corresponding sample/chromosome box.


Chromosome
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
Primary tumors
K-Ras(G12D) p53 WT
2240a
2240b
2347a
2062a
K-Ras(G12D) p53 +/-
1202c
1267a
1247d
K-Ras(G12D) p53 R270H/+
1265c
1532a
NT1a
Cell lines*
LKR10
LKR13
+C3
-D1, -D3
-A1
+A1
-A4, -F1
-A1
-A1.2 to A1.3
-D1
+qA3.3
+A1
-A1
-A1
-H4
+B
+A1, +A2
-A1.2, +A3
-A1.3
+C1
+B3.1
+C4
+A1
whole chromosome gain
whole chromosome loss
sub-chromosomal gain
sub-chromosomal loss

alteration was chromosome 12 gain, which was found in 5/10 tumors. Other recurrent changes in order of prevalence are chromosomes 19 gain (3/10), 3 gain (2/10), 16 gain (2/10), 9 loss (2/10), and 11 loss (2/10). Interestingly, among the 28 tumors analyzed, one tumor 1247d showed copy number changes in multiple focal regions. Representative data are shown in Figure 4.

p53 loss or mutation has been shown to accelerate lung tumor progression in the *Kras* mouse model. When compared to the lungs of $Kras^{G12D}$;*p53*+/- mice, $Kras^{G12D}$;$p53^{R270H/+}$ lungs showed an increase in tumor number and an increased proportion of higher grade tumors (Jackson et al. 2005). Because p53 functions in multiple pathways that maintain genomic integrity, we questioned whether the differences in lung tumorigenesis kinetics in $Kras^{G12D}$;*p53*+/+, $Kras^{G12D}$;*p53*+/- and $Kras^{G12D}$;$p53^{R270H/+}$ mice can be explained by elevated genomic instability due to p53 loss or mutation. In our study, it appeared that histological grading best correlated with the presence or absence of genetic changes, and that germline p53 genotype of the mice did not affect the spectra of lesions significantly. Tumors that showed any changes were of grade 2 or higher, with one exception, NT1a, which contained a mix of grade 1 and 2 characteristics. The genotypes of mice that gave rise to the tumors did not appear to affect this trend for the most part. The one note was while all grade 2-3 tumors in $Kras^{G12D}$;*p53*+/+ and $Kras^{G12D}$;*p53*+/- mice exhibited ROMA changes, two lesions of comparable grade $Kras^{G12D}$;$p53^{R270H/+}$ xsmice (1291a, 1782a) were silent. It remains unclear whether this silence was due to the difference in genotype, or a result of varying

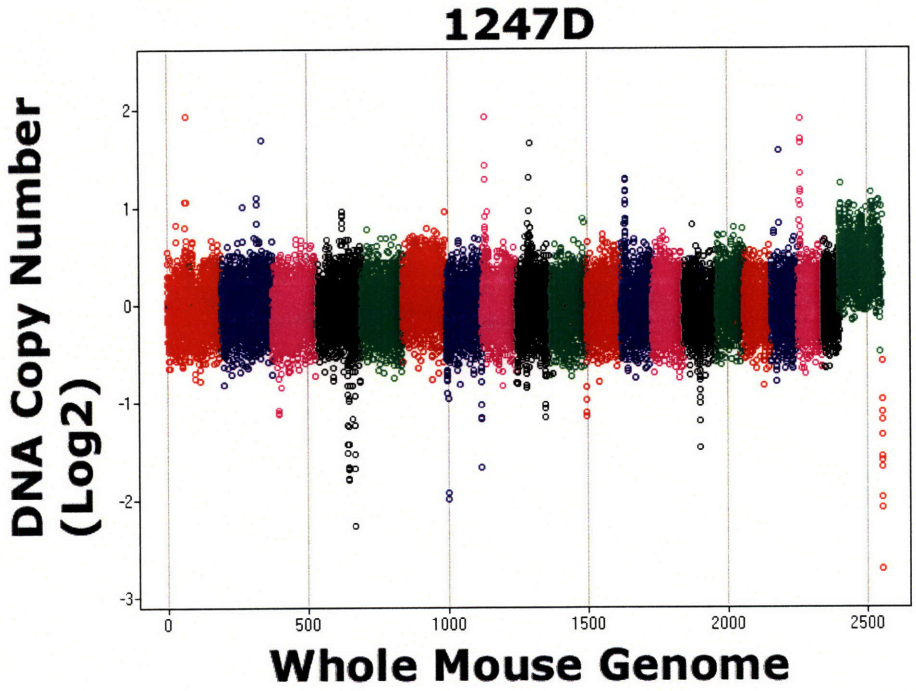
Figure 4: Representative ROMA moving median plots of lung tumors

A moving median plot shows Cy5 to Cy3 signal ratios of from labeled tumor vs. tail DNA. The Y-axis is the log10 fluoresence ratio and the X-axis is an index of the probes genomic order, based on UCSC mouse May 2004 (mm5) annotated assembly. Data from same chromosome are labeled with same color.

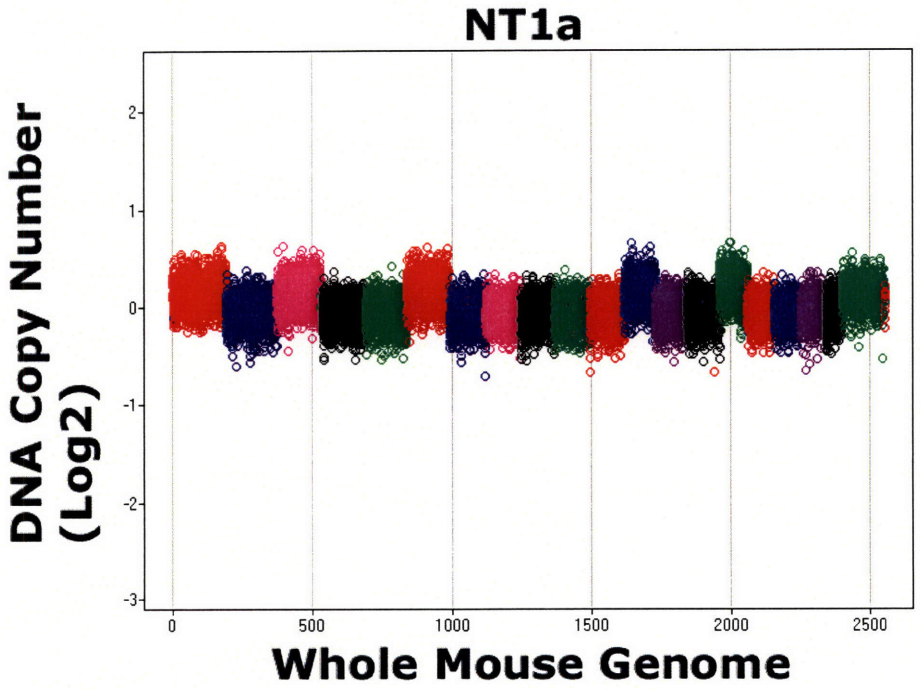
A) Moving median plot of 1247d tumor genome, representative of samples showing sub-chromosomal gains and losses.

B) Moving median plot of NT1a tumor genome, representative of samples showing whole chromosomal gains and losses.

A


1247D
DNA Copy Number (Log2)
2
1
0
-1
-2
-3
0
500
1000
1500
2000
2500
Whole Mouse Genome

B


NT1a
DNA Copy Number (Log2)
2
1
0
-1
-2
-3
0
500
1000
1500
2000
2500
Whole Mouse Genome

stromal contamination in tumors that might have muted ROMA signals. Human subjectivity in histological grading should also be considered.

Detection of focal changes in mouse lung cancer cell lines

In addition to the 28 primary lung tumors, we performed ROMA to analyze copy number changes in two mouse lung cancer cell lines, LKR10 and LKR13. The cell lines were derived from the tumor-bearing lungs of one LA1 mouse, which carried a latent allele of $Kras^{G12D}$ that got spontaneously activated by recombination (Johnson et al. 2001). In DNA from both cell lines, ROMA detected increased copies of whole chromosomes 6 and 19, which also showed recurrent gains in the primary samples. In addition, ROMA revealed multiple focal changes in DNA copy number in both cell lines that were unseen in most of the primary tumors (Figure 3). The sizes of lesions range from 0.095 to 3.9Mb, each containing one or more gene or EST. The changes were nearly identical in the two cell lines, suggesting a common clonal origin. Apparent phenotypic differences of these two cell lines were likely due to smaller-size genetic changes missed by ROMA and/or epigenetic differences.

Common deletion of Csmd1 in mouse and human lung cancer cell lines

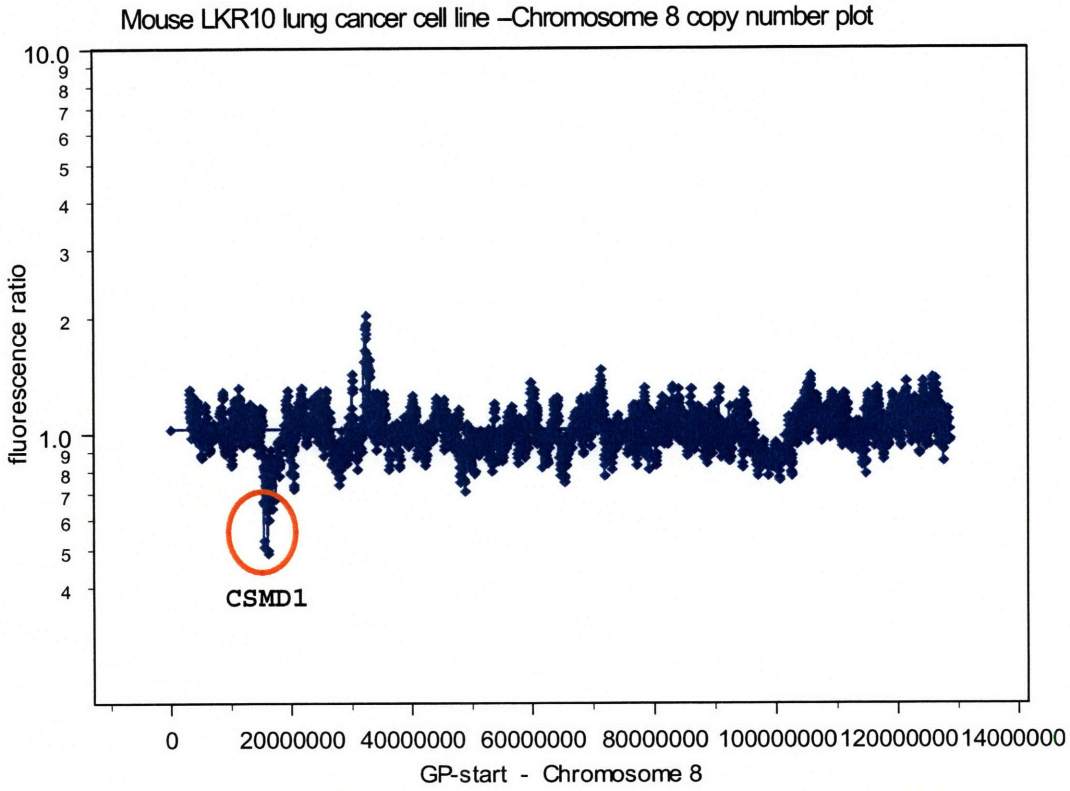
We compared data from our mouse lung tumor set to ROMA data from a human lung cancer cell line H460 (David Mu, unpublished data). Similar to the mouse LKR10 and 13 cells lines, H460 contains multiple focal copy number alterations. Syntenic regions containing changes were compared. Intriguingly, the orthologs of CUB and sushi multiple domains 1 (*Csmd1*) gene was reduced in copy number in mouse LKR10/LKR13 cell lines (Figure 5B) and human H460 cell line (Figure 5B). As one of the biggest genes,

Figure 5: Csmd1 deletion in mouse and human lung cancer cell lines

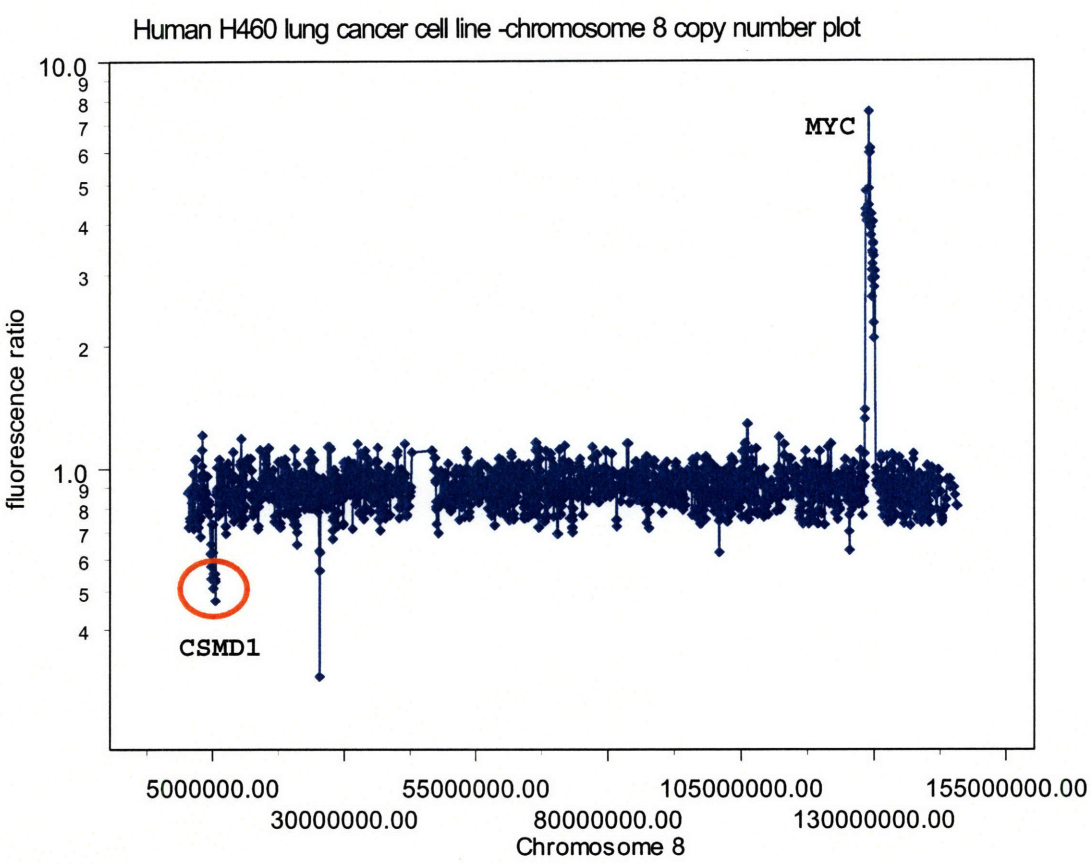
A) ROMA moving median plot of chromosome 8 in mouse lung cancer cell line LKR10. A ~2Mb focal deletion was seen on chromosome 8: 15321241-17425932 (UCSC mouse May 04 assembly). Csmd1 is the only known RefSeq gene present.

B) ROMA moving median plot of chromosome 8 of human lung cancer cell line H460. A focal lesion was seen in chr8: 4743548-5613707 (UCSC human April 03 assembly), within an 8p region that is conserved in synteny with mouse chromosome 8. Csmd1 is also the only known RefSeq gene present.

A


Mouse LKR10 lung cancer cell line –Chromosome 8 copy number plot
fluorescence ratio
10.0
1.0
CSMD1
0
20000000
40000000
60000000
80000000
100000000
120000000
14000000
GP-start - Chromosome 8

B


Human H460 lung cancer cell line -chromosome 8 copy number plot
fluorescence ratio
10.0
1.0
MYC
CSMD1
5000000.00
30000000.00
55000000.00
80000000.00
105000000.00
130000000.00
155000000.00
Chromosome 8

Csmd1 covers 1.64Mb in the mouse genome and 2.06Mb in human. Loss of human chromosome 8p, where *Csmd1* resides, is a common event observed in ~15% of lung cancers and lung cancer cell lines (David Mu, unpublished data). The identity of the critical tumor suppressor gene(s) in the region is still unclear. Csmd1 is a potential candidate, which has been found deleted or inactivated in cancer or tumor cell lines (Scholnick and Richter 2003). We attempted to compare Csmd1 expression level in normal lung vs. lung tumors in datasets generated using gene expression microarrays (Alice Shaw, unpublished). However, the absence of transcript signal in normal lung made comparison difficult.

Discussion

In this study, we tested the application of the mouse ROMA platform to characterize genetic alterations in retinoblastomas and lung adenocarcinomas in different genetically engineered mouse models. We were able to detect various forms of copy number changes including amplifications, deletions, and chromosomal gains and losses. We can hereby compare our results to some of the common genetic changes in human cancers. Future analysis using larger number of tumors via this strategy can thus be very informative.

On the technical capabilities of mouse ROMA platform

Array CGH analysis compares relative DNA sequence copy number between genomes. The value of a platform depends largely on its spatial resolution. The detectable changes in this study range from 95kb to whole chromosomal gains and losses. Among the sub-Mb lesions detected and subsequently verified was an amplification of the *N-Myc*

oncogene in metastatic retinoblastomas, proving the usefulness of the mouse ROMA platform. Several different kinds of array CGH platforms are available for genome-wide studies in both mouse and humans. Arrays using BACs as probe elements are highly sensitive but spatial resolution is limited by the size of BACs, which range from 150kb to 200kb (Pinkel and Albertson 2005). Short oligonucleotides can greatly reduce the limit, despite its lower sensitivity needs to be compensated by the use of a higher amount of DNA in hybridization and the averaging of signals from 3-5 adjacent probes to make a reliable call (Pinkel and Albertson 2005). ROMA starts with a digestion-amplification protocol from as little as 50ng of DNA to make representations, which reduce hybridization noise by lowering genomic complexity (Lucito et al. 2003). Thus, while ROMA requires little starting materials as needed for BAC arrays, it also takes advantage of a sub-Mb resolution with an oligonucleotide array platform. This ability of the mouse ROMA was demonstrated by the detection of N-Myc amplification in retinoblastomas.

The availability of both human and mouse ROMAs has opened a new avenue for comparative genomic analyses. In many human cancers, large-scale chromosomal abnormalities are common but delimitation of critical genetic regions is often difficult. There are 300+ syntenic segments covering over 90% of the mouse and human genomes. These conserved regions have been evolutionarily rearranged within and between chromosomes. The shuffling has made the mouse genome useful in assessing the relative importance of various syntenic segments that correspond to a large human chromosomal area relevant in cancer. Indeed, CGH screens on mouse cancers have been performed to delimit regions of genetic importance in the human disease (Hodgson et al. 2001; Hackett

et al. 2003). Continual development of mouse CGH platforms like ROMA in parallel to human ones would enable more comparative studies to be done.

On the genetics of retinoblastomas

The Rb/p130 DKO mouse model developed tumors that histologically resemble human retinoblastoma (MacPherson et al. submitted). ROMA was used to assess if the similarity is also present at the genetic level. CGH and cytogenetic studies have suggested the majority of human retinoblastomas contained chromosomal imbalance. Chromosomes 1q and 6p gains are the most frequent, found in over 50% of all human retinoblastomas (Mairal et al. 2000; Chen et al. 2001; Lillington et al. 2003; Zielinski et al. 2005). In the ROMA analysis of eight mouse retinoblastomas from Rb/p130 DKO mice, gain of chromosomes 1 and 12 were seen in half of the samples. Interestingly, mouse chromosome 1 has three syntenic blocks on human chromosome 1q: 1q23.2-32, 1q32.1, and 1q32.2-42.1. Extra copies of the same orthologous gene(s) in one or more of these regions might be selected for in both mouse and human retinoblastomas. Some studies suggested the minimal region of gain in human to be 1q31. Further experiments are needed to determine if this is the case in our mouse model.

Mouse chromosome 12, also apparently gained in half of our tumor DNA samples, has a region syntenic to human chromosome 2p. Furthermore, within chromosome 12, a minimally overlapping region of 136kb was seen amplified with even higher magnitude. The only known gene residing in this region is *N-Myc*. A common childhood nervous system tumor, neuroblastoma, frequently has *N-Myc* amplification, which marks rapid tumor progression (Brodeur et al. 1984). *N-Myc* overexpression has also been implicated

in other neuronal cancers including human retinoblastomas (Mairal et al. 2000). While the amplification is often associated with gains of other genes on human chromosome 2p (Mairal et al. 2000; Lillington et al. 2003; Zielinski et al. 2005), *N-Myc* appears to be the critical gene in our mouse model. Our sample set contains only 8 late-stage metastatic tumors, making an extended study necessary to establish the timing of the *N-Myc* amplification.

On the genetics of lung adenocarcinomas

Our lab has described mouse models of lung cancer based on an expression of an activated $Kras^{G12D}$ allele from its endogenous locus (Jackson et al. 2005). Microarray-based gene expression analyses have been performed to assess the molecular similarity between human lung cancer and tumors from a *Kras*-initiated mouse model (Sweet-Cordero et al. 2005). At the genetic level, one CGH study was previously done with a 2K BAC array platform on a set of *Kras*-induced lung tumors (Sweet-Cordero et al. 2006). Among the 59 tumors analyzed in that experiment, recurrent whole-chromosomal changes were detected but no focal copy number gains or losses could be seen. The current study described in this chapter differed from the prior one in three major ways: (1) a new technical platform was employed. With 84K arrayed oligonucleotides and a protocol to enhance signal to noise ratios using genomic representations, ROMA has the potential to provide an enhanced resolution to reveal focal changes that might be present. (2) This study has encompassed tumors from mice with double *Kras* and *p53* mutations in the sample set, in order to assess if germline *p53* status affects level or spectrum of genetic alterations. (3) The current study analyzed tumors from inbred mice with a pure 129S4/svJae background, instead of mice from a C57BL6J x 129S4/SvJae F1 cross. It

has been suggested that genetic background can affect types of lesions in tumors. Of particular note is that mitotic recombination can be suppressed in F1 hybrids from two different parental strains.

Here, ROMA was used to analyze 28 Kras-induced mouse lung tumors and 2 tumor-derived cell lines. In summary, we observed recurrent gains of chromosomes 3, 6, 12, 19 and losses of 9 and 11 in tumors that were graded 2 or higher using criteria set by Jackson et al. 2005. Focal subchromosomal copy number changes were detected in one tumor and the two LKR cell lines.

p53 and genetic changes in Kras-initiated lung cancer model

Despite a p53 germline lesion leads to more total tumors and more histologically advanced tumors in the *Kras*-initiated lung cancer model, there was no striking difference in mutation spectra of histologically comparable tumors from these mice. *p53* can play a role in inducing cell cycle arrest, senescence, and apoptosis in face of DNA damage. Its loss can promote chromosomal instability and tumor progression in various mouse models (e.g.(Hingorani et al. 2005)). In our study, a *p53* mutation appears to mainly act by providing a more permissive environment for the outgrowth of *Kras* mutant cells, instead of altering the kinds of cooperative genetic elements *Kras* needs to drive tumor progression.

Whole chromosomal copy changes

Changes in whole chromosome copy number constitute the overwhelming majority of changes seen in the primary lung tumor samples. This suggests non-disjunction was a

major driver to create secondary lesions for *Kras*-initiated lung tumors to progress beyond grade 2 in histology. The presence of chromosomal copy number changes in almost all grade 2+ tumors has implied a selection of other genetic changes in the tumor initiating Kras mutant cells. The recurrent whole chromosomal copy number changes suggested gain or loss of one or more genes on these chromosomes might be important in this model of lung cancer.

Chromosome 6 gain was seen in 80% of tumors with alterations found by ROMA. Mouse chromosome 6 harbors the Kras gene, pulmonary adenoma susceptibility 1 (Pas1) locus, and contains a region in synteny to a human 3q segment, all of which have been implicated in lung cancer. It will be of particular interest to test whether the chromosome 6 gain consistently corresponds to a copy increase of the mutant *Kras* allele. The *Kras* gene resides at the distal arm of chromosome 6. In vitro transformation of rodent cells can be promoted by amplification of the mutant *ras* gene (Sorrentino et al. 1988). One particularly intriguing experiment involved a study of Rat-1 cells engineered to express an *H-ras* activating mutation from its endogenous locus. Cells heterozygous for the mutation underwent spontaneous transformation at a low frequency, and most transformed cells had the mutant allele amplified (Finney and Bishop 1993). Finally, *Kras* amplification in human lung carcinoma has also been observed (Pulciani et al. 1985). On the other hand, other loci on chromosome 6 might have been selected in our Kras-initiated tumors. For instance, *Pas1* is a quantitative trait locus that affects lung cancer predisposition in mice. While *Kras* is the primary candidate, polymophisms in other genes such as *Las1* and *Lrmp* have also been associated with lung cancer

susceptibility(Manenti et al. 2004). In addition, human chromosome 3q has been reported to gain in copy in various subtypes of lung cancer including adenocarcinomas (Testa et al. 1994; Pei et al. 2001; Balsara and Testa 2002; Garnis et al. 2006). Orthologous genes within the syntenic regions might be important for tumorigenesis in both species.

Chromosome 12 gain was the second most frequent alteration, observed in 50% of all analyzed samples containing any changes. Interestingly, both mouse chromosomes 12 and 6 have regions syntenic to human chromosome 7. Furthermore, the conserved areas corresponding to the two mouse chromosomes tend to cluster adjacent to each other in two major areas on chromosome 7. Polysomy of human chromosome 7, as well as regional amplifications in both 7p and 7q arms have been seen in cytogenetic and array CGH studies on lung tumors and tumor cell lines (Balsara and Testa 2002; Wong et al. 2003; Kim et al. 2005; Garnis et al. 2006). On 7p, one report has described 7p22.1-22.3 and 7p11.2-15.3 gains in over 80% of samples (Garnis et al. 2006). Coincidently, one cluster with syntenic conservation to chromosome 6 and 12 happens to be within 7p14-22, which contains a many known genes including the developmentally important *HoxA* gene cluster and beta-integrin 8. On the other arm 7q, areas with synteny include 7q22-36.1, which harbors the T-cell recepter-beta gene cluster among others.

Among the chromosomal losses, chromosomes 9 and 11 were each reduced in copy in two tumors. A distal part of chromosome 9 is syntenic to human chromosome 3p21-22. Loss of human chromosome 3p is the most common event observed in lung cancer (Zabarovsky et al. 2002). In particular 3p21 loss is observed as one of the earliest event,

which can be detected in the pre-malignant epithelium of smokers. As for chromosome 11, its distal arm has syntenic conservation to the entire human chromosome 17, where p53 resides.

Sub-chromosomal copy changes

Despite being the minority, one tumor and the two LKR cell lines exhibited multiple focal copy number changes. Some of the detected lesions were below 1 Mb in sizes. The two LKR cell lines shared almost identical lesions, suggesting they have the same clonal origin. A cell line represents a subclone within a tumor mass that got selected to expand in tissue culture. Different sets of genetic criteria were likely required for during cell line establishment vs. clonal outgrowth in vivo, which may explain the difference in kinds of lesions observed. In addition, according to the clonal evolution model (Nowell 1976), each tumor mass is likely a composite of heterogeneous subclones, which have individual proliferation rates and fates under selection. The absence of focal ROMA signals in the primary tumors does not directly imply the absence of focal lesions within its subclones. Instead, alterations may mask each other inside a heterogeneous tumor. Data from primary tumors are essentially an average of all the differences. In addition, ROMA lacks the ability to detect general polyploidy or non-reciprocal translocations, both of which will appear as constant total tumor to normal DNA ratios in all types of CGH studies.

Detectable focal changes can allow delineation of critical genetic elements. When we compared ROMA datasets from the mouse LKR lung cancer cell lines to a human lung cancer cell line, focal deletions involving the *Csmd1* mouse and human orthologs were observed. While the physical loss of this locus still remains to be validated, ROMA

seemingly has the resolving power to reveal common orthologous changes down to a single-gene level in a cross-species study. Future comparative genomic studies using ROMA or other array CGH platforms will likely provide more molecular insights on tumorigenesis.

Materials and Methods

Lung tumor DNA isolation

All mouse protocols were approved by the animal care committees at the Massachusetts Institute of Technology. *KRas*^{LSLG12D}, *KRas*^{LSLG12D}; *p53*+/- and *KRas*^{LSLG12D}; *p53* ^{R270H}/+ mice on a 129S4/SvJae background was infected with adenovirus Cre as described in Jackson et al. 2005. Lung tumors were dissected from the lungs of mice 22-24 weeks after infection. One portion of each tumor was fixed in formalin, sectioned in paraffin, and stained in hematoxylin and eosin. Histological grading of each tumor was assigned based on a 1-5 scale as described in Jackson et al.. Remaining tumor material was stored at -80°C prior to DNA isolation. Tail from each mouse was collected for use as normal control. DNA was extracted from thawed tissues using reagents and protocols in Puregene DNA isolation kit (Gentra Systems, Inc.). LKR10 and LKR13 mouse lung cancer cell lines were derived from K-Ras LA1 mouse on a 129S4/SvJae background.

Retinoblastoma tumor generation

As described in (MacPherson et al. submitted)), Pax6 a-enhancer Cre mice were bred with *Rb*$^{lox/lox}$; *p130-/-* animals. Mice were maintained on a mixed C57BL/6; 129SvJ;

FvB/n background. Late stage metatastic tumors were collected from mice at time of sacrifice, 183±30 days of age. Samples were frozen at -80°C until DNA isolation.

ROMA analysis

Genomic DNA from tumors and tails of corresponding mice were paired for each experiment. DNA was digested with BglII to obtain low complexity representations (LCRs) of the genomes. LCRs from tumor and normal tissues were differentially labeled with Cy 5 and Cy3 respectively by random priming. The hybridizations to oligonucleotide microarrays were performed as described in (Lucito et al. 2003).The design of the mouse ROMA arrays is described elsewhere (Lucito et al., in preparation). Array images were acquired with an Axon GenePix 4000B scanner. The raw array data were globally normalized. A moving window of a 5 data-points was used to smoothen the raw data by assigning the median value of the moving window to each central data-point.

Southern analysis

Genomic DNA was digested with *EcoR*I. *N-Myc* probe was a 1.1kB cDNA fragment including sequence from exons 2 and 3 that was obtained by PCR of a mouse embryo brain cDNA library using the following primers: 5'gaggacagcgcagataaagg and 5' cctcactcctaatccggtc. *Rosa26* probe hyrbidization to chromosome 6 was used as control and performed as described in (Soriano 1999). Southern blotting was performed using standard protocols. Blot was first probed with the *N-Myc* probe, stripped and rehybridized to the *Rosa26* probe to control for loading discrepancies. Band intensities on phosphoimager scans were quantified using ImageQuant.

Acknowledgements

The authors would like to thank Denise Crowley and Mike Brown for histology preparation. We also thank Rob Bronson for pathology consultation. D. Mu is supported by a grant from Joan's Legacy Foundation. T.J. is a Howard Hughes Investigator.

References

Balsara, B.R. and Testa, J.R. 2002. Chromosomal imbalances in human lung cancer. *Oncogene* **21**(45): 6877-6883.

Barrett, M.T., Scheffer, A., Ben-Dor, A., Sampas, N., Lipson, D., Kincaid, R., Tsang, P., Curry, B., Baird, K., Meltzer, P.S., Yakhini, Z., Bruhn, L., and Laderman, S. 2004. Comparative genomic hybridization using oligonucleotide microarrays and total genomic DNA. *Proc Natl Acad Sci U S A* **101**(51): 17765-17770.

Brodeur, G.M., Seeger, R.C., Schwab, M., Varmus, H.E., and Bishop, J.M. 1984. Amplification of N-myc in untreated human neuroblastomas correlates with advanced disease stage. *Science* **224**(4653): 1121-1124.

Chen, D., Gallie, B.L., and Squire, J.A. 2001. Minimal regions of chromosomal imbalance in retinoblastoma detected by comparative genomic hybridization. *Cancer Genet Cytogenet* **129**(1): 57-63.

Dhami, P., Coffey, A.J., Abbs, S., Vermeesch, J.R., Dumanski, J.P., Woodward, K.J., Andrews, R.M., Langford, C., and Vetrie, D. 2005. Exon array CGH: detection of copy-number changes at the resolution of individual exons in the human genome. *Am J Hum Genet* **76**(5): 750-762.

Finney, R.E. and Bishop, J.M. 1993. Predisposition to neoplastic transformation caused by gene replacement of H-ras1. *Science* **260**(5113): 1524-1527.

Garnis, C., Lockwood, W.W., Vucic, E., Ge, Y., Girard, L., Minna, J.D., Gazdar, A.F., Lam, S., Macaulay, C., and Lam, W.L. 2006. High resolution analysis of non-

small cell lung cancer cell lines by whole genome tiling path array CGH. *Int J Cancer* **118**(6): 1556-1564.

Hackett, C.S., Hodgson, J.G., Law, M.E., Fridlyand, J., Osoegawa, K., de Jong, P.J., Nowak, N.J., Pinkel, D., Albertson, D.G., Jain, A., Jenkins, R., Gray, J.W., and Weiss, W.A. 2003. Genome-wide array CGH analysis of murine neuroblastoma reveals distinct genomic aberrations which parallel those in human tumors. *Cancer Res* **63**(17): 5266-5273.

Hingorani, S.R., Wang, L., Multani, A.S., Combs, C., Deramaudt, T.B., Hruban, R.H., Rustgi, A.K., Chang, S., and Tuveson, D.A. 2005. Trp53R172H and KrasG12D cooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice. *Cancer Cell* **7**(5): 469-483.

Hodgson, G., Hager, J.H., Volik, S., Hariono, S., Wernick, M., Moore, D., Nowak, N., 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**(4): 459-464.

Jackson, E.L., Olive, K.P., Tuveson, D.A., Bronson, R., Crowley, D., Brown, M., and Jacks, T. 2005. The differential effects of mutant p53 alleles on advanced murine lung cancer. *Cancer Res* **65**(22): 10280-10288.

Jobanputra, V., Sebat, J., Troge, J., Chung, W., Anyane-Yeboa, K., Wigler, M., and Warburton, D. 2005. Application of ROMA (representational oligonucleotide microarray analysis) to patients with cytogenetic rearrangements. *Genet Med* **7**(2): 111-118.

Johnson, L., Mercer, K., Greenbaum, D., Bronson, R.T., Crowley, D., Tuveson, D.A., and Jacks, T. 2001. Somatic activation of the K-ras oncogene causes early onset lung cancer in mice. *Nature* **410**(6832): 1111-1116.

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**(5083): 818-821.

Kim, T.M., Yim, S.H., Lee, J.S., Kwon, M.S., Ryu, J.W., Kang, H.M., Fiegler, H., Carter, N.P., and Chung, Y.J. 2005. Genome-wide screening of genomic alterations and their clinicopathologic implications in non-small cell lung cancers. *Clin Cancer Res* **11**(23): 8235-8242.

Lillington, D.M., Kingston, J.E., Coen, P.G., Price, E., Hungerford, J., Domizio, P., Young, B.D., and Onadim, Z. 2003. Comparative genomic hybridization of 49 primary retinoblastoma tumors identifies chromosomal regions associated with histopathology, progression, and patient outcome. *Genes Chromosomes Cancer* **36**(2): 121-128.

Lucito, R., Healy, J., Alexander, J., Reiner, A., Esposito, D., Chi, M., Rodgers, L., Brady, A., Sebat, J., Troge, J., West, J.A., Rostan, S., Nguyen, K.C., Powers, S., Ye, K.Q., Olshen, A., Venkatraman, E., Norton, L., and Wigler, M. 2003. Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation. *Genome Res* **13**(10): 2291-2305.

MacPherson, D., Conkrite, K., Tam, M., Mukai, S., Mu, D., and Jacks, T. submitted. Murine bilateral retinoblastoma exhibiting rapid onset, metastatic progression and N-myc gene amplification.

Mairal, A., Pinglier, E., Gilbert, E., Peter, M., Validire, P., Desjardins, L., Doz, F., Aurias, A., and Couturier, J. 2000. Detection of chromosome imbalances in retinoblastoma by parallel karyotype and CGH analyses. *Genes Chromosomes Cancer* **28**(4): 370-379.

Manenti, G., Galbiati, F., Gianni-Barrera, R., Pettinicchio, A., Acevedo, A., and Dragani, T.A. 2004. Haplotype sharing suggests that a genomic segment containing six genes accounts for the pulmonary adenoma susceptibility 1 (Pas1) locus activity in mice. *Oncogene* **23**(25): 4495-4504.

Nowell, P.C. 1976. The clonal evolution of tumor cell populations. *Science* **194**(4260): 23-28.

Pei, J., Balsara, B.R., Li, W., Litwin, S., Gabrielson, E., Feder, M., Jen, J., and Testa, J.R. 2001. Genomic imbalances in human lung adenocarcinomas and squamous cell carcinomas. *Genes Chromosomes Cancer* **31**(3): 282-287.

Pinkel, D. and Albertson, D.G. 2005. Array comparative genomic hybridization and its applications in cancer. *Nat Genet* **37 Suppl**: S11-17.

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. *Nat Genet* **20**(2): 207-211.

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. *Nat Genet* **23**(1): 41-46.

Pulciani, S., Santos, E., Long, L.K., Sorrentino, V., and Barbacid, M. 1985. ras gene Amplification and malignant transformation. *Mol Cell Biol* **5**(10): 2836-2841.

Scholnick, S.B. and Richter, T.M. 2003. The role of CSMD1 in head and neck carcinogenesis. *Genes Chromosomes Cancer* **38**(3): 281-283.

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**(4): 399-407.

Soriano, P. 1999. Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nat Genet* **21**(1): 70-71.

Sorrentino, V., McKinney, M.D., Drozdoff, V., Hume, C.R., and Fleissner, E. 1988. Spontaneous or carcinogen-mediated amplification of a mutated ras gene promotes neoplastic transformation. *Oncogene Res* **2**(2): 189-195.

Sweet-Cordero, A., Mukherjee, S., Subramanian, A., You, H., Roix, J.J., Ladd-Acosta, C., Mesirov, J., Golub, T.R., and Jacks, T. 2005. An oncogenic KRAS2 expression signature identified by cross-species gene-expression analysis. *Nat Genet* **37**(1): 48-55.

Sweet-Cordero, A., Tseng, G.C., You, H., Douglass, M., Huey, B., Albertson, D., and Jacks, T. 2006. Comparison of gene expression and DNA copy number changes in a murine model of lung cancer. *Genes Chromosomes Cancer* **45**(4): 338-348.

Testa, J.R., Siegfried, J.M., Liu, Z., Hunt, J.D., Feder, M.M., Litwin, S., Zhou, J.Y., Taguchi, T., and Keller, S.M. 1994. Cytogenetic analysis of 63 non-small cell lung carcinomas: recurrent chromosome alterations amid frequent and widespread genomic upheaval. *Genes Chromosomes Cancer* **11**(3): 178-194.

Wong, M.P., Fung, L.F., Wang, E., Chow, W.S., Chiu, S.W., Lam, W.K., Ho, K.K., Ma, E.S., Wan, T.S., and Chung, L.P. 2003. Chromosomal aberrations of primary lung adenocarcinomas in nonsmokers. *Cancer* **97**(5): 1263-1270.

Zabarovsky, E.R., Lerman, M.I., and Minna, J.D. 2002. Tumor suppressor genes on chromosome 3p involved in the pathogenesis of lung and other cancers. *Oncogene* **21**(45): 6915-6935.

Zielinski, B., Gratias, S., Toedt, G., Mendrzyk, F., Stange, D.E., Radlwimmer, B., Lohmann, D.R., and Lichter, P. 2005. Detection of chromosomal imbalances in retinoblastoma by matrix-based comparative genomic hybridization. *Genes Chromosomes Cancer* **43**(3): 294-301.

Chapter 3

A multiplexed SNaPshot™ protocol for genome-wide mouse SNP genotyping and its application to detect loss-of-heterozygosity in murine lung tumors

Mandy Tam[1], Alejandro Sweet-Cordero[1], Brian Weiss[2], David Ginzinger[3], Julie Weng[3], Annie Fung[3], and Tyler Jacks[1,4]

[1]Massachusetts Institute of Technology, Department of Biology and Center for Cancer Research, 77 Massachusetts Avenue, Cambridge, MA 02139 USA

[2]Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, OH 45229, USA

[3]University of California at San Francisco, Comprehensive Cancer Center, San Francisco, CA94143, USA

[4]Howard Hughes Medical Institute, 4000 James Bridge Road, Chevy Chase, MD 20185, USA

The author, A.S.C. and B.W shared the initial work to validate individual SNPs. The author designed the multiplexing protocol, obtained lung tumor DNA, performed all PCR and genotyping data anlaysis. SNaPshot™ reactions were performed by J.W. and A.F..

Abstract

While the choice of human SNP genotyping methodologies is broad, few have been applied for whole-genome analysis in the mouse. In this chapter, we presented a new and validated protocol. We identified 358 published SNPs individually for polymorphism in 129S4/svJae vs. C57BL/6J strains and worked out a protocol to genotype a panel of 147 markers in the mouse using the SNaPshot™ (Applied Biosystems) genotyping system. Our method uses a standard DNA sequencing machine to resolve single-base extension (SBE) products of genotyped SNPs, enabling easy adaptation by most standard laboratories. We tested the use of the assay as a mapping tool of LOH in mouse tumors. We analyzed 20 lung tumor DNA samples from a *Kras*-driven mouse model of lung cancer and detected several LOH markers including loss of wild-type p53.

Introduction

Genetic alterations are the underlying cause of cancer. Cancer-causing genes can be generally dividing into two categories: 'proto-oncogenes', which when activated by mutation or overexpression can promote abnormal proliferation, growth, and differential, and 'tumor suppressor genes', which when inactivated lose their normal regulatory roles in these processes. In the classical 'two-hit' hypothesis, loss of heterozygosity (LOH) is believed to promote cancer by inactivating tumor suppressor gene function (Knudson 1971). Genome-wide LOH screens by allelotyping are useful tools to identify loci that may harbor tumor suppressors genes. (Sokolov 1990)

Mouse tumor models are valuable experimental tools to study human cancer genetics. In many human cancers, large-scale chromosomal abnormalities are common but finding of the critical regions is often difficult. Through evolutionary shuffling, most individual chromosomes in one mammalian species can be delimited into regions of synteny that are conserved in other species on separate chromosomes. This is true between human and mouse. Given such syntenic structure and that >90% of the human genome is covered by conserved areas in the mouse (Waterston et al. 2002), LOH studies of mouse cancer may provide an invaluable perspective to delineate large human genetic lesions.

Genome-wide LOH screens have been performed on several mouse models of human cancer (Dietrich et al. 1994; Radany et al. 1997; Herzog et al. 2002; Wu et al. 2002; Benavides et al. 2003). LOH screens are traditionally performed with simple sequence length polymorphism (SSLP) markers. Genotyping of SSLP using PCR can be performed

by most laboratories without extensive capital investment but is cumbersome. Single nucleotide polymorphism (SNP) is becoming the marker of choice for various purposes of genome-wide screening, due to its dense genome coverage and the availability of several high-throughput genotyping methods. In the mouse, genome-wide SNP screens have been performed to study haplotype structure (Wade et al. 2002; Wiltshire et al. 2003; Frazer et al. 2004; Zhang et al. 2005), delineate strain relationships (Petkov et al. 2004b; Pletcher et al. 2004), and facilitate genetic mapping (Grupe et al. 2001; Pletcher et al. 2004; Owens et al. 2005; Moran et al. 2006). In the screening of LOH regions in mouse cancer, however, no report using SNPs as markers has been made to date.

Various methods of SNP genotyping have been recently developed (Syvanen 2001). However, the cost of specialized instruments such as mass spectrometry machines may set a barrier for performing SNP mapping studies. This chapter presents a multiplexed protocol for genome-wide mouse SNP genotyping in mice using SNaPshot™(Applied Biosystems), which couples single-base extension (SBE) reactions to capillary electrophoresis and fluorescence detection. The use of only standard PCR and sequencing machines should make the assay adaptable for use by most laboratories. We tested the protocol and the principle of using SNPs to identify LOH regions with a mouse lung cancer model. Mice expressing a $Kras^{G12D}$ allele develops tumors that histologically resemble human non-small cell lung cancer (NSCLC) (Johnson et al. 2001) and share similar gene expression profiles (Sweet-Cordero et al. 2005). We sought to characterize genetic changes associated with these tumors. We observed several cases of whole-

chromosomal LOH, suggesting other genetic factors can cooperate with a *Kras* mutation in tumorigenesis in this model.

Results

Validation of SNPs in 129S4/SvJae vs. C57BL/6J strains

358 SNPs distinguishing C57BL/6J and 129/Sv or 129S1/SvImJ sub-strains were selected from databases originating from two independent large-scale SNP discovery studies (Germer et al. 2000; Lindblad-Toh et al. 2000b). We designed SBE assays to individually test the genotypes of these chosen SNPs in C57BL/6J (B6) vs.129S4/SvJae (129S4) DNA. Supplemental Table 1 lists the SNP genotyping results of DNA from B6 and 129S4 mice maintained in-house and those from B6x129S4 F1 hybrids. The B6 and 129S4 DNA used for this analysis did not come from true parental mice crossed to generate the tested F1 hybrids. By our assay, a total of 172 SNPs were confirmed to be polymorphic in B6 and129S4, as defined by observed heterozygosity in F1 and homozygosity for different alleles between the parental strains. Among the unconfirmed markers, 41 were found to be non-polymorphic, 31 appeared 'heterozygous' in 129S4, 31 looked 'heterozygous' in B6 (10 of which also 'heterozygous' in 129), 67 were inconsistent among the three tested genotypes, and 25 failed genotyping. Our testing of a different 129 sub-strain vs. B6 may explain why some markers are non-polymorphic. Alternatively, the non-polymorphic markers may be false positives from the large-scale discovery screens. The high number of 'heterozygous' markers in DNA from the inbred strains was surprising. Cross hybridization of an SBE primer to another genomic position

may produce two allelic peaks and a 'heterozygous' score. It is also possible that despite extensive inbreeding, certain loci on the parental strains remain polymorphic, which may result in the inconsistent genotypes between tested F1 hybrids and the non-biological parental B6 and 129S4 strains inbred in-house. Nevertheless, 172 markers across the mouse genome were individually validated to be polymorphic between the two tested strains.

Establishment of a protocol for genome-wide multiplex SNP typing in the mouse

147 validated SNPs were chosen for genotyping in a multiplex fashion (Figure 1). The positional distribution of SNPs in the mouse genome is illustrated in Figure 2. The mean distance between markers was 14.5Mbp, and at least 2 SNPs were present in each of 19 autosomes (Table 1). A single-base extension method was used to genotype the SNP panel. Schematic of the procedure is shown in Figure 1. The protocol employed 93 PCRs to amplify the 147 SNP targets (Supplemental Table 2). Pooling of PCR products reduced the number of subsequent SBE reactions to 15. The amounts of individual PCR products in each pool were adjusted according to the strength of the PCRs. In each SBE reaction, 5-11 SNPs were genotyped with primers of varied lengths, resolvable by capillary electrophoresis (Table 2). SBE primer concentration was adjusted to correct for differential signal strength. A representative SBE reaction output is shown in Figure 3. The relative height of the two allelic peaks from each marker was quantified on the electrograph.

Figure 1: Schematic for SNP genotyping

A multiplex strategy was used to genotype 147 SNPs in the mouse genome. (A) 93 PCRs were used to amplify genomic DNA containing 1-2 SNP targets in each reaction. Each SNP is depicted as [allele 1/allele 2]. (B) PCR products were pooled as templates for each single-base extension (SBE) reaction, in which fluorescently labeled dideoxynucleotide triphosphates were incorporated into primers of different lengths. 15 SBE reactions each genotyping 5-11 SNPs were performed. (C) SBE products were resolved on an ABI 3700 capillary DNA sequencer.

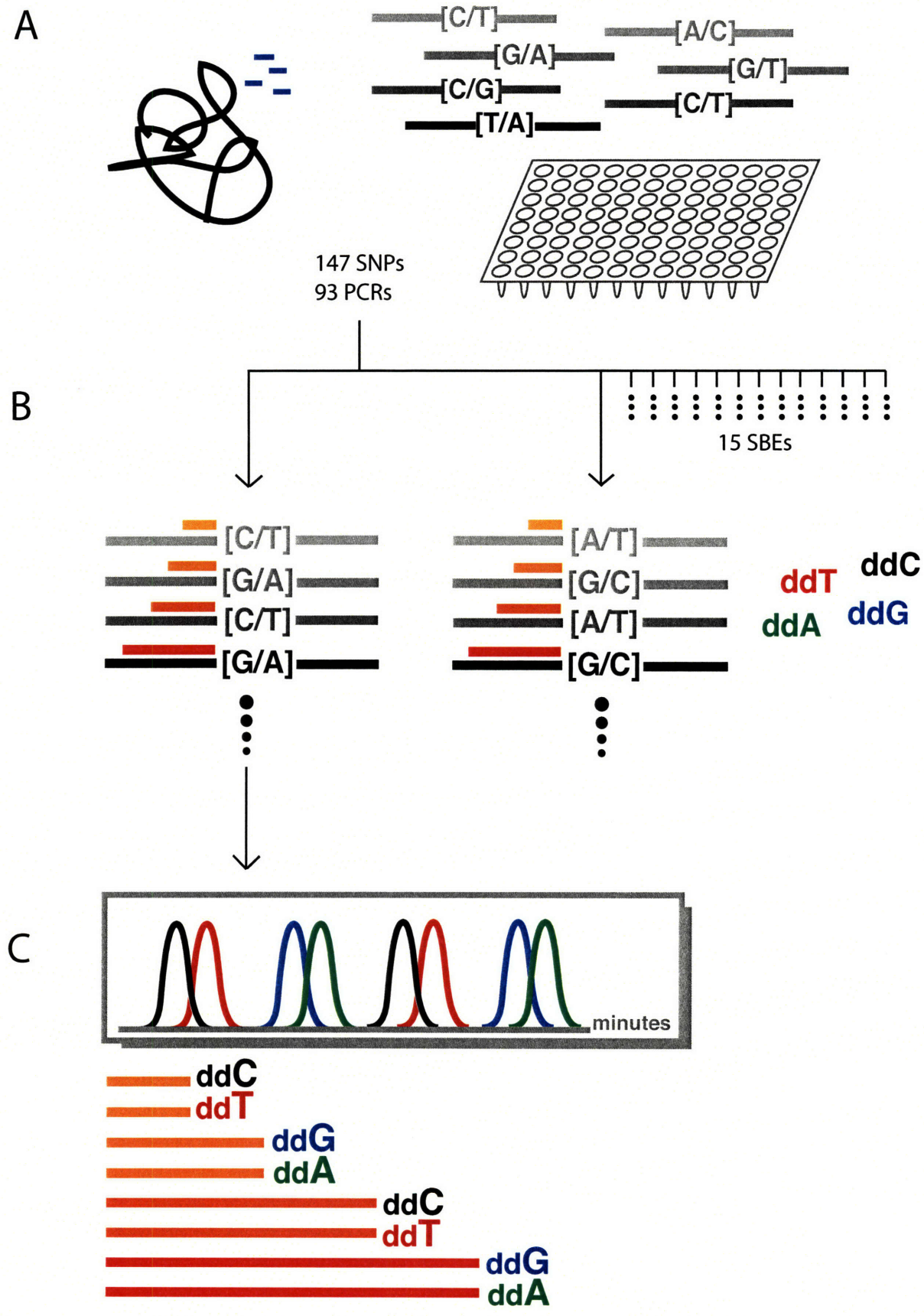

A
[C/T]
[G/A]
[C/G]
[T/A]
[A/C]
[G/T]
[C/T]
147 SNPs
93 PCRs
B
15 SBEs
[C/T]
[G/A]
[C/T]
[G/A]
[A/T]
[G/C]
[A/T]
[G/C]
ddT
ddC
ddA
ddG
C
minutes
ddC
ddT
ddG
ddA
ddC
ddT
ddG
ddA

Figure 2: Positional representation of 147 screening markers in mouse genome

Positions of markers on chromosomes are based on UCSC mouse Feb 2006 (mm8) assembly

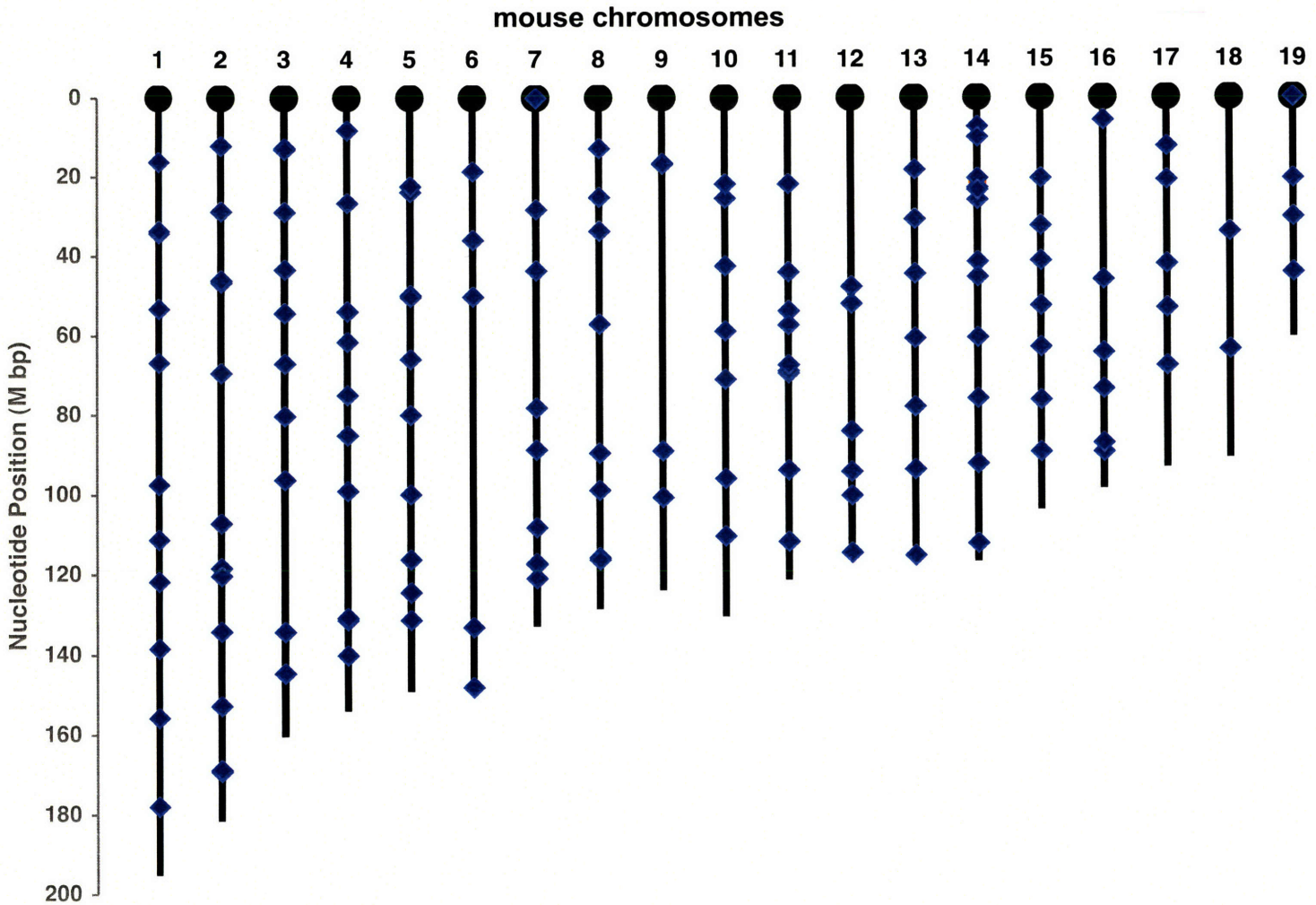

mouse chromosomes
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
Nucleotide Position (M bp)
0
20
40
60
80
100
120
140
160
180
200

Table 1: Chromosomal distribution of SNP markers in the genome-wide screen

Chromosome	Marker count	Minimum distance (Mb)	Maximum distance (Mb)	Mean distance (Mb)
1	11	0.50	30.80	16.18
2	13	0.00	37.85	13.10
3	10	0.47	38.10	14.69
4	10	0.56	31.77	14.69
5	10	0.33	26.04	12.11
6	5	14.26	82.94	32.38
7	9	0.26	34.48	15.13
8	9	0.04	32.49	12.97
9	4	0.29	72.25	28.09
10	7	3.65	24.98	14.78
11	10	0.06	24.41	10.02
12	6	4.37	32.01	13.40
13	7	12.61	21.65	16.23
14	12	0.54	20.18	9.57
15	7	8.73	13.39	11.50
16	6	2.26	40.36	16.76
17	5	8.54	21.19	13.84
18	2	29.59	29.59	29.59
19	4	9.77	20.52	14.74
Whole Genome	**147**			**14.54**

Table 2: List of multiplexed SBE assays for whole-genome SNP genotyping

SBE primer sequence and genotype of each SNP in 129xB6 F1 hybrid are shown. SBE primers were designed and organized such that those typing alternative nucleotide pairs differ in at least 2 bases of length.

Multiplex SBE group	Assay ID	RefSNP ID	SBE primer sequence	Primer Length	SNP
1A	412	rs3089257	GCCATTCGTCCCAGGG	16	CT
	439	rs3023183	TCTCTGTGCCCACAGCCA	18	GA
	471	rs3022975	GCTCTGGGAATGTGCTTTTC	20	CT
	883	rs3668662	AAAGGCAGTGGGTACACATCAT	22	GA
	772	rs3704164	GCTGCCATATGAAGATCTCCTCTA	24	CT
	405	rs3022887	TTTTGGGTGTTTCTATGATAACGCTC	26	GA
	779	rs3660209	GTCAATAGGTGAGAAAAATATCAGACTG	28	CT
	786	rs3720966	TTTTTTTTCACCCACGTTGTGACTTTAAGC	30	GA
	830	rs3714631	TTTTTTTTTTTTGAGGTGTGTTGGTGTGCACA	32	CT
	850	rs3090435	TTTTTTTTTTTGGTACTATCCCTGGCTTTTCACA	34	GA
	426	rs3090833	TTTTTTTTTTTGAAAAAATTGCTTGTACTGTAGTCT	36	CT
1B	109	rs3023468	GCACATGGTTCTGCCACA	18	GC
	452	rs3023449	TGACACAACGCTGTTCCAGG	20	TA
	868	rs3710059	TTTTGGAGCTGGCACATCCACT	22	GC
	455	rs3023436	TTTTCCAGGGAGTAAAACATCAGG	24	TA
	879	rs3695889	TTTTTCAAGTTAGAAGCATTGCCCTC	26	GC
	823	rs3713224	TTTTACCAGTTTCATTTCCCATACTTCT	28	AT
	490	rs3090645	TTTTTTTGGCTACTACACTAGCAAATCCATAG	32	TA
	922	unmapped	TTTTTTTTTTTAGAGCCAGACATAGTAGCACACGGA	36	AT
	893	rs3688361	TTTTTTTTCCTTAGCAGTTTAGGAATATTTAGATAGTTAA	40	AT
1C	808	rs3669262	TTATCAGCACCCGTCCCA	18	GT
	415	rs3022979	TGAGGGCATTGGTTTCTTTC	20	CA
	821	rs3704980	GACTAGTCCACATTGGTGAGCA	22	GT
	461	rs3023258	CCTCCCATTATTTTTTTTTCTGAG	24	CA
	422	rs3023045	TTAACTTGGAATATCAGGCTTTCTTT	26	GT
	79	rs3089102	TTTTTTTGAGTGTCACATGGATTTGCAC	28	CA
	600	rs3022802	TTTTTCAAAGGGTATGAGAATATGGACTGG	30	GT
	840	rs3694308	TTTTTTCAAGTATATGGACTTGGAAGACAATG	32	CA
	453	rs3023456	TTTTTTTTTATGTGCCTTCTTATGTCGACTTTAT	34	GT
	895	rs3662097	TTTTTTTTTTTTATTGGACTCTATATTTTGTCTGGGCT	38	GT
1D	761	rs3665023	TGGCTGAGGGACTTGTGC	18	CT
	478	rs3022989	ACAAAACTGCCAGTTGCTTC	20	GA
	33	rs3023057	TTTTATCGAGGTGCCTTTTGTC	22	CT
	489	rs3090908	CAGAAACATAATTTCAAAGTTGCA	24	GA
	757	rs3682376	CCTTTCTCCTATCTTTACTTCCTAGC	26	CT
	803	rs3683689	TTTTTTCACTGTGGTCAGACAGAATGCA	28	GA
	713	rs3684370	TTTTTTTTTTCAGCAAGGAGTGTGTATGCA	30	CT
	81	rs3023379	TTTTATATTTATCTCACTGTGAAGTCTGCCTA	32	GA
	831	rs3676476	TTTTTTTTTTTTCACTCAGTAGGTCAGAGCAGGG	34	CT
	892	rs3717068	TTTTTTTTTTTTAGTCTACTTTCAGTGCTGTCCCAT	36	GA
1E	790	rs3685393	CAGTTCCAAGCACCCACA	18	CT
	762	rs3679837	ACTGCCAGTTCATGACCTCC	20	GA
	741	rs3696551	CTGTCCCCGTAGACTAGACCTT	22	CT
	770	rs4137954	GCTCTGGTAAGTTTAACACACTCC	24	GA
	756	rs3688884	TTTTTTTGATGGGTGGGTGTGTCAGT	26	CT
	700	rs3659426	TTTTTTTTGGAGTCAGACAGGAATGGAA	28	GA
	51	rs3023194	TTTTTTTTTTTTTGCCTTTAGGTTTCATGC	30	CT
	913	rs3677860	TTTTTTTTTTCTGGCTGTTCATTATTTGTGCA	32	GA
	701	rs3681957	TTTTTTTCTCAGGTACTGAGTGAGTTCCTAGACT	34	CT
	911	rs3697014	TTTTTTTTTTTGCTTAAGCTTGTGTATCAACTGCCT	36	GA
	812	rs3657504	TTTTTTTTTTTTTCAAGACCACATCTCCTATTCCTTCT	38	CT
1F	797	rs3696966	TAGGAGGGAACGGAGGC	17	GA
	746	rs3657668	GCGGCTTCATTCTCCATCT	19	CT
	882	rs3674239	TCTTCCACTCTTTGGTCCTCC	21	GA
	727	rs3681675	TCAGGTAATGGAAAATCACTCAG	23	CT
	839	rs3667625	CAGACATTCTTACTCCATCATCTCC	25	GA
	768	rs3716232	TTTTTTTACACATCACAAGGCCACCTA	27	CT
	809	rs3724533	TTTTTTGGTAGGCAGGTACCAGAATCTCA	29	GA
	769	rs3694785	TTTTTTTATACGACTTAGCTACAGTCCCTGG	31	CT
	857	rs3705482	TTTTTTAATATATATCCAGTGGAATTGAGTGGT	33	GA
	898	rs3667466	TTTTTTTTTTTTCAGGGACACACACATATTTTGCT	35	CT
1G	912	rs3702150	CAGGCGGGCTAAATTCA	17	CT
	763	rs3686956	ATGTGACAGGGAAGTTGGC	19	GA
	742	rs3722968	TCTCTCCTAGCTCATCCCATG	21	CT
	814	rs3708958	GCTGAGTCACGGTACATAAAGTTGT	25	CT
	833	rs3726717	TTTTTCAGAAGCTCAGAAAGCATCAAG	27	GA
	844	rs3691937	TTTTTTTTTGCCTCTGTTGTGGCTAATCA	29	CT
	843	rs3657720	TTTTTTTTCCTTGTAATAAGCCACAGCATCT	31	GA
	827	rs3090608	TTTTTTTTTTTTTCCGAGTGCTGACTCTGGGTT	33	CT
	817	rs3664582	TTTTTAAGAAGTGTCCCAAATCCTTTCTATATAGT	35	GA
	834	rs3689513	TTTTTTTTTTTTGCTTTGGAGACATATTGTGTGGTTA	37	CT

1H	856	rs3697769	CCTCACATGTGGACTCAGGC	20	GT
	766	rs3724779	AGAGGACCTAATTGAGGACTGC	22	CA
	86	rs3089070	TTTCCTAGTTATTTCTGTGGTCTT	24	GT
	75	rs3023347	ACATCCATATTTACAAGGTCATAGAA	26	CA
	715	rs3704392	GTTATTTATGCTCCAACAGTTATTGAAA	28	GT
	788	rs3713871	TTTTTTTTCCCTTAGCTTTCAAGTCCTTGC	30	CA
	436	rs3089474	TTTTTTGTTTAGAATGTTGCTCTTAAGGGTTT	32	GT
	835	rs3669413	TTTTTTTCAAACTTGTGGTCGTGATAGATATTAG	34	CA
	465	rs3088800	TAACAACTTTAGTTAACTAGAAATACTAAGTCTTGA	36	GT
2A	446	rs3090586	TAGCGCACAGTGCCAGAA	18	CT
	4	rs3022839	GAGTAACATCACAGCCTTCG	20	GA
	32	rs3023037	TTTATGGTGCCAGAAAATCAAC	22	CT
	488	rs3090260	TTAAAGTCTCAACTCCATCTTTCC	24	GA
	721	rs3664018	TTTTTTTCACAGCCTCAGAAAGTCCC	26	CT
	218	rs3023026	TTTTTCCACACCTCCACTATTATAAAGC	28	GA
	820	rs3708255	TTTTTTTTTTTTCCTGTCCCTTCACCAGGG	30	CT
	859	rs3658370	TTTTTTTTTTCAATGTGGGTAAAACTGGCAAT	32	GA
	96	rs3023416	TTTTTTCCCTTCAGAAATGAAAATTAATCTACTA	34	CT
	80	rs3023386	TTTTTTTTTAAGTTCATCATTCCCTAGGATGTTATA	36	GA
	445	rs3089912	TTTTTTTTTTTTTTTGAAAAAGGCAGTGACAAAGTATG	38	CT
2B	818	rs3722942	GCTGCCCATTCTCACCT	17	GA
	722	rs3681847	AGGATTGGATCAGCCATCA	19	CT
	858	rs3674616	ATGCCCAGAGAGTGATCTAGAAG	23	CT
	764	rs3711535	TTTTTTTCCATTTGTTCCAGAGGCA	25	GA
	339	rs3023161	CACCACTTGATATAGGAATGTACACTT	27	CT
	811	rs3023409	TTGTGATATGTGGAAGTTATATAAGCTTC	29	GA
	732	rs3706063	TTTTTTTTTTCATGTGGCTGGTTACCTCTC	31	CT
	118	rs3023175	TTTTTTTTTTTGCACAAGTCAGCATCAACGCAT	33	GA
	822	rs3685188	TTTTTTTTTTTCTTCTGTTCTGCTGGTTACAGCTC	35	CT
2C	724	rs3664805	AAGATCATCAGGGGCCTG	18	CT
	20	rs3022960	TTGAAACATGGAGACAAGGC	20	GA
	448	rs3023243	TTTCGGCAACTGACTTTGGACA	22	CT
	483	rs3088822	GTACCTGTGAGTATTCAGTCAGCA	24	GA
	67	rs3023265	TTCCTTATTTTGCCAGTCTCCTTACT	26	CT
	800	rs3671678	TTTTTTTTTGTGGTGGGGGAGCAATATG	28	GA
	321	rs3021908	TTTTTTTTTTTGACGGTGTGTGCCTACAAT	30	CT
	860	rs3701351	TTTTTTTTGCCAAACAATTGTAAGAATGTGTG	32	GA
	826	rs3663534	TTTTTTTTTTTTCCAAACTCCAAGTTCTTCAGCC	34	CT
	201	rs3023256	TTTTTTTTTTTTTTTCTGGGAGAGGTAACTGCTAACT	36	GA
	890	rs3658201	TTTTTTTTTTTTGGTTTTGATTCTTCAGTGTAGTTTGG	38	CT
2D	787	rs3712403	CCTTCCTGTCTGTTCCAGC	19	AT
	794	rs3654982	TCTAGTCCACCAGCAGCAGAAAC	23	AT
	845	rs3023067	TTTTTCTTGCTGTCCTTTTGAGCTGAG	27	TA
	854	rs3710192	TTCTCCATGTTAAGCATTACAATTATGACTA	31	AT
	851	rs3723894	TTTTTTGGCTAGTTCATAGAGTATCAGAAATGTGT	35	AT
2E	767	rs3672332	ACAGTGAGGAGGGGCCA	17	CT
	435	rs3090731	CAGGTCTCCAGCTGAAAGC	19	GA
	726	rs3667376	GTTTGTGACCTGGTCTCTGTG	21	CT
	750	rs3709317	CTGCAATGAACATCACAGAGC	21	GA
	740	rs3666032	GTGAAGGACAGACAGACAGACAG	23	GA
	738	rs3711350	TGACAATTTCTCACATGGTATTAGATC	27	GA
	731	rs3674631	TTTTTTTTTCAGGAACGCAGTGATTGTTC	29	CT
	862	rs3700023	TTTTTTTTGATTGCCCTATAGCCATTACCTG	31	GA
	832	rs3669022	TTTTTTTTTTTTCTTAGCCCTCCCCAACTTACC	33	CT
	606	rs3023117	CAATAGATATTTAGATGTTGCTATTGTTTATCTAC	35	GA
	829	rs3660910	TTTTTTTTTTTAATTTTGCTTTATGTGGCTGTCTATC	37	CT
2F	909	rs3721297	AGGCAGGTCCATGCAGG	17	CA
	484	rs3089436	CACATGGGGACTGTCCAAA	19	GT
	841	rs3707288	TCACCTGCTCGTATTCCTGGA	21	CA
	801	rs3726430	GGGTAGGGGTAGGAAGTAGAGAG	23	GT
	486	rs3090719	TTTTGGTCCCACCTTGTTACAGGTC	25	CA
	745	rs3713298	GGTAGCTGCTAAATTAATCTTCAAGAG	27	GT
	733	rs3672323	TTTCTTAACTGTGAAGAACTAAACTGCAG	29	CA
	101	rs3090912	TTTTTTTAAGTACTGATGGCTTTGAGTCTTA	31	GT
	789	rs3713838	TTTTTTTTTTTGAGGGTCAGAGCACTTGCAGTA	33	CA
2G	900	rs3716435	CCCACGTTCCAACACACA	18	CT
	730	rs3666331	CCTCAGTGAACTGCACATCC	20	GA
	714	rs3662163	GCTGTATAAACTATGCCCCCAA	22	CT
	37	rs3023051	TTTTGAGAATGAAATGAACACCAG	24	GA
	463	rs3088501	ATGCTAGTAGGAAGACTCTGGAACTA	26	CT
	413	rs3090381	TTAAGTACTTGGGTATGAAGTTCTCAAA	28	GA
	723	rs4137557	TTTTTTTTTTTGGACCATTCTCCGTGTTCT	30	CT
	816	rs3726591	TTTTTTTTTTTGGGAGGTCGGTATTAGGAGAC	32	GA
	748	rs3685067	TTTTTTTTTTTTAGATGCCTCATCTGATAACAGG	34	CT
	828	rs3706262	TTTTTTTTTTTTTCTCTGACTCCAGTCCCTCTGGG	36	GA
	799	rs3719410	TTTTTTTTTTTCAATGAGCTTTGAATTTCTGCTAATAA	38	CT

Implementation of SBE assay to identify LOH regions in mouse lung tumors

20 lung tumor DNA samples from 129S4xB6 F1 mice were subjected to LOH screening using the genome-wide SNP panel. The tumors range from grade 1-3 using criteria described in Jackson et al. (Table 3). 3828 out of 4290 SNP data points (89%) could be scored with confidence. For each data point, allelic imbalance factor (AIF) was calculated as described in materials and methods. The cutoff for positive call was set at AIF of >3 or <-3. With the assumption that most loci in the normal tissue of each F1 mouse are heterozygous, normal DNA was not subjected to the whole-genome screening protocol. Instead, assays were repeated on tumor and the corresponding tail DNA on individual markers that exhibited LOH in the first-pass screen. Figure 3 illustrates the global LOH landscape observed in this first pass. Markers scored positively for LOH are listed in Table 4. The majority (>98%) of tested loci remained heterozygous in the lung tumors, although certain regions of LOH were suggested by 44 positive markers. B6 markers on chromosome 14 were lost in DNA from one grade 3 tumor (840a), which was obtained from a $Kras^{LA2}$ mouse (Figure 4, Table 4). Two other grade 3 tumors were present in the sample set, 866h and 870a from $Kras^{LA2}$; *p53+/-* mice. Both tumors appeared to have lost all 129S4 markers on chromosome 11 (Figure 4, Table 4). 870a also showed LOH along chromosomes 9 and 10. In the initial screen, 16/44 positively scored markers were focal, defined by the absence of LOH in adjacent markers (Table 4). However, when 8 of these markers were screened in the second round, none had a positive score. Remaining markers still need to be screened in second round. On the other hand, positive scoring was concordant for most markers on chromosome 14 in 840a tumor DNA (Table 4).

Figure 3: Representative assay for multiple SNPs in single SBE reaction

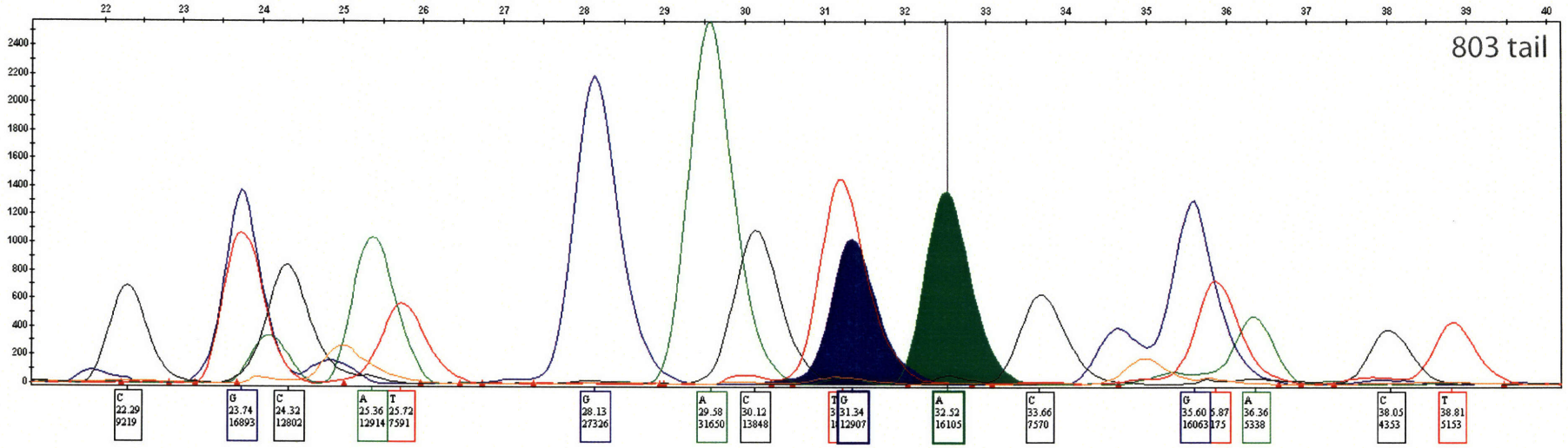
11 SNPs were genotyped simultaneously in the SBE group 1A (Table 2).

A) Electrographic data from genotyping of SBE group 1A on 803tail DNA. The X-axis is electrophoretic shift, a function affected by SBE primer length and dye chemistry of labeled ddNTPs. The Y-axis is signal amplitude reflecting the allele quantity. The blue, black, red, and green peaks correspond to the nucleotides G, C, T, A respectively. The G and A peaks of the SNP marker 786 are highlighted.

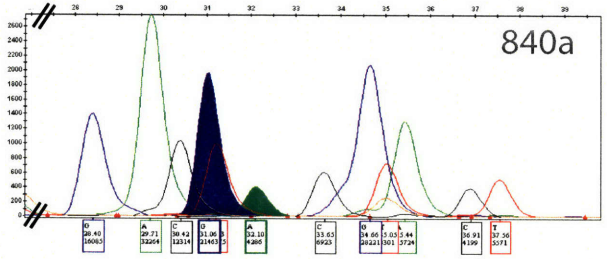
B) Electrographic output from genotyping of SBE group 1A on 840a tumor DNA. Signal ratio of G to A peaks of marker 786 is altered, suggesting an LOH event.

C) Summary of genotyping results of all group 1A markers for 803tail and 840a. Signal ratio of the two allelic peaks was quantified for each SNP. AIF was calculated after normalizing peak ratios of sample to averages from three normal controls (see materials and methods).

A



B



C

Assay ID	Marker	Sample	Genotype		Peak Positions		Peak Heights		Height Ratio	AIF
439	M-05763_2	803tail	n/a		n/a		n/a		n/a	n/a
		840a	G	A	20.89	22.5	377	641	0.5881	1.0689
412	M-11390_1	803tail	C	T	22.29	23.74	697	1071	0.6508	0.6731
		840a	C	T	23.22	24.93	289	1440	0.2007	2.1828
883	WI_WGS_17_20370149	803tail	G	A	23.74	25.36	1373	1041	1.3189	0.9145
		840a	G	A	24.34	26.35	1644	695	2.3655	1.6401
471	M-07408_2	803tail	C	T	24.32	25.72	840	566	1.4841	1.1809
		840a	C	T	24.8	26.17	570	641	0.8892	0.7076
772	WI_WGS_10_41002641	803tail	n/a		n/a		n/a		n/a	n/a
		840a	n/a		n/a		n/a		n/a	n/a
405	M-08889_1	803tail	G	A	28.13	29.58	2187	2573	0.8500	0.9946
		840a	G	A	28.4	29.71	1399	2767	0.5056	0.5917
779	WI_WGS_10_108120320	803tail	C	T	30.12	31.23	1096	1455	0.7533	0.9392
		840a	C	T	30.42	31.23	1037	1003	1.0339	0.6843
786	WI_WGS_14_7229989	803tail	G	A	31.34	32.52	1033	1371	0.7535	0.8334
		840a	G	A	31.06	32.1	1959	406	4.8251	5.3368
830	WI_WGS_2_152763270	803tail	C	T	33.66	35.87	634	735	0.8626	0.8381
		840a	C	T	33.65	35.05	614	724	0.8481	0.8240
850	D6Mit201_2	803tail	G	A	35.6	36.36	1304	475	2.7453	0.7292
		840a	G	A	34.66	35.44	2072	1304	1.5890	1.2599
426	M-02221_2	803tail	C	T	38.05	38.81	381	437	0.8719	1.0441
		840a	C	T	36.91	37.56	393	510	0.7706	0.9229

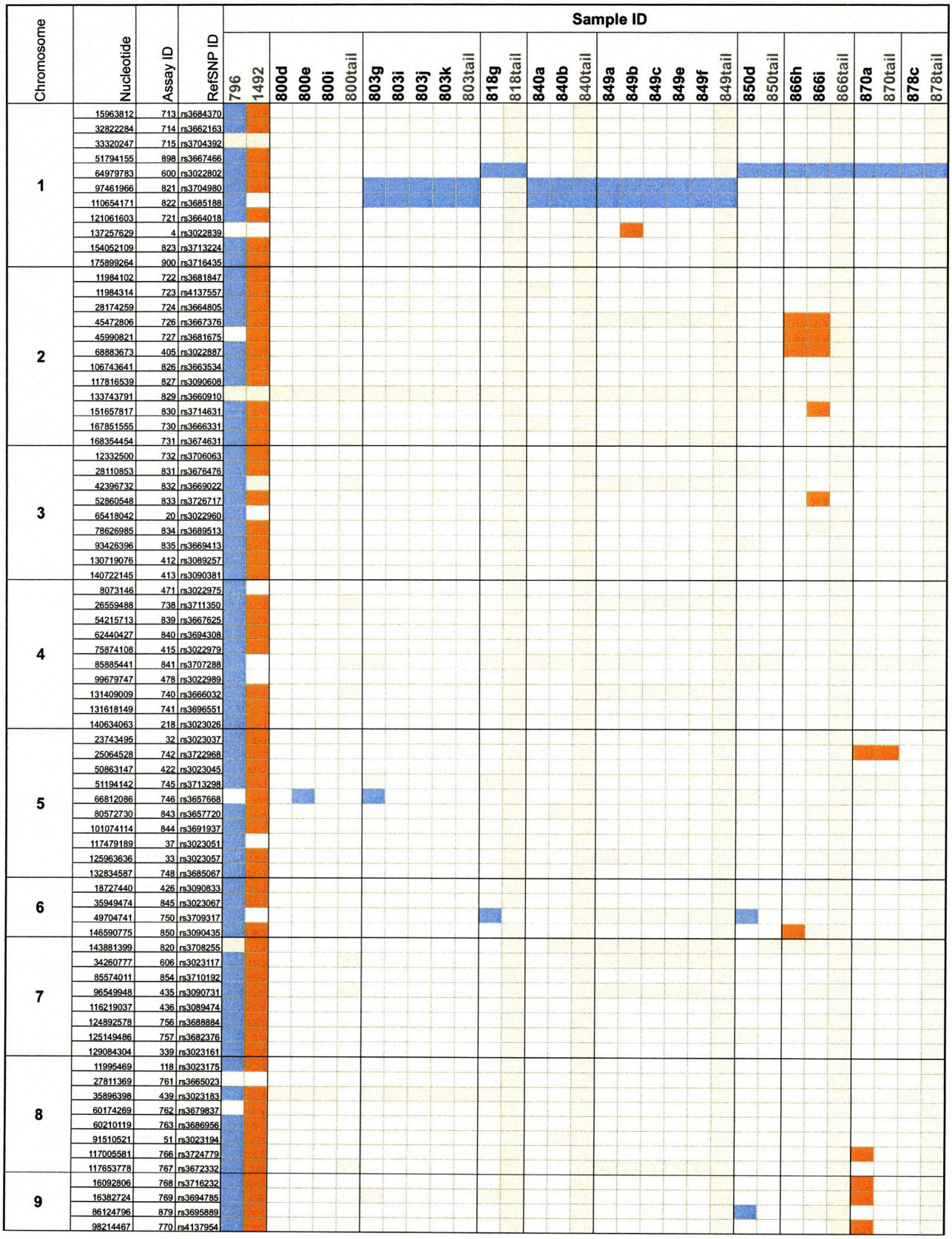
Table 3: Characteristics of primary lung tumors analyzed

The SNP screen was passed on 20 lung tumors from 129xB6 F1 hybrid mice and control tail DNA from two parental strains. The parental strains were not the biological parents of the F1 mice. The table lists the genotype of the mice from which individual lung tumors were dissected out. Histological grading was assigned using a grade 1-5 scale as described previously (Jackson et al. 2005).

Sample ID	Mouse genotype	Histology
796	WT (C57CL6J)	(tail DNA)
1492	*Kras*^{LA2} (129S4/SvJAe)	(tail DNA)
800d	*Kras*^{LA2}	Grade 1
800e	*Kras*^{LA2}	Grade 2+
800i	*Kras*^{LA2}	Grade 1-2
803g	*Kras*^{LA2}	Grade 2,
803i	*Kras*^{LA2}	Grade 1
803j	*Kras*^{LA2}	Grade 1
803k	*Kras*^{LA2}	Grade 2,
818g	*Kras*^{LA2}	Grade 2,
840a	*Kras*^{LA2}	Grade 2+ (25%), grade 3 (75%)
840b	*Kras*^{LA2}	Grade 2,
849a	*Kras*^{LA2}	Grade 1
849b	*Kras*^{LA2}	Grade 1
849c	*Kras*^{LA2}	Grade 1
849e	*Kras*^{LA2}	Grade 1
849f	*Kras*^{LA2}	Grade 1
850d	*Kras*^{LA2}	Grade 2+
866h	*Kras*^{LA2}; *p53*+/-	Grade 2 (10%), 90% grade 3-3+ (90%)
866i	*Kras*^{LA2}; *p53*+/-	Grade 2, bone in tumors
870a	*Kras*^{LA2}; *p53*+/-	Grade 3+
878c	*Kras*^{LA2}; *p53*$^{R270H/+}$	Grade 1

Figure 4: Genome-wide LOH screening of primary mouse lung adenocarcinomas

Grid illustrating data from the first-pass SNP screen: columns represent different samples; rows are markers organized by the chromosomal positions from UCSC Feb 2006 (mm8) assembly. Markers with positive AIF score are colored as follows: orange=AIF>3 (loss of B6 allele), blue=AIF<0.3 (loss of 129S4 allele). Shaded boxes are non-informative markers. DNA from 20 primary tumors, 2 parental strains, and representative F1 tails were screened on a genome-wide scale. Samples labeled with the same numerical prefix in the IDs were collected from the same mouse.


Chromosome
Nucleotide
Assay ID
RefSNP ID
Sample ID
796
1492
800d
800e
800i
800tail
803g
803i
803j
803k
803tail
818g
818tail
840a
840b
840tail
849a
849b
849c
849e
849f
849tail
850d
850tail
866h
866i
866tail
870a
870tail
878c
878tail
1
15963812 713 rs3684370
32822284 714 rs3662163
33320247 715 rs3704392
51794155 898 rs3667466
64979783 600 rs3022802
97461966 821 rs3704980
110654171 822 rs3685188
121061603 721 rs3664018
137257629 4 rs3022839
154052109 823 rs3713224
175899264 900 rs3716435
2
11984102 722 rs3681847
11984314 723 rs4137557
28174259 724 rs3664805
45472806 726 rs3667376
45990821 727 rs3681675
68883673 405 rs3022887
106743641 826 rs3663534
117816539 827 rs3090608
133743791 829 rs3660910
151657817 830 rs3714631
167851555 730 rs3666331
168354454 731 rs3674631
3
12332500 732 rs3706063
28110853 831 rs3676476
42396732 832 rs3669022
52860548 833 rs3726717
65418042 20 rs3022960
78626985 834 rs3689513
93426396 835 rs3669413
130719076 412 rs3089257
140722145 413 rs3090381
4
8073146 471 rs3022975
26559488 738 rs3711350
54215713 839 rs3667625
62440427 840 rs3694308
75874108 415 rs3022979
85885441 841 rs3707288
99679747 478 rs3022989
131409009 740 rs3666032
131618149 741 rs3696551
140634063 218 rs3023026
5
23743495 32 rs3023037
25064528 742 rs3722968
50863147 422 rs3023045
51194142 745 rs3713298
66812086 746 rs3657668
80572730 843 rs3657720
101074114 844 rs3691937
117479189 37 rs3023051
125963636 33 rs3023057
132834587 748 rs3685067
6
18727440 426 rs3090833
35949474 845 rs3023067
49704741 750 rs3709317
146590775 850 rs3090435
7
143881399 820 rs3708255
34260777 606 rs3023117
85574011 854 rs3710192
96549948 435 rs3090731
116219037 436 rs3089474
124892578 756 rs3688884
125149486 757 rs3682376
129084304 339 rs3023161
8
11995469 118 rs3023175
27811369 761 rs3665023
35896398 439 rs3023183
60174269 762 rs3679837
60210119 763 rs3686956
91510521 51 rs3023194
117005581 766 rs3724779
117653778 767 rs3672332
9
16092806 768 rs3716232
16382724 769 rs3694785
86124796 879 rs3695889
98214467 770 rs4137954

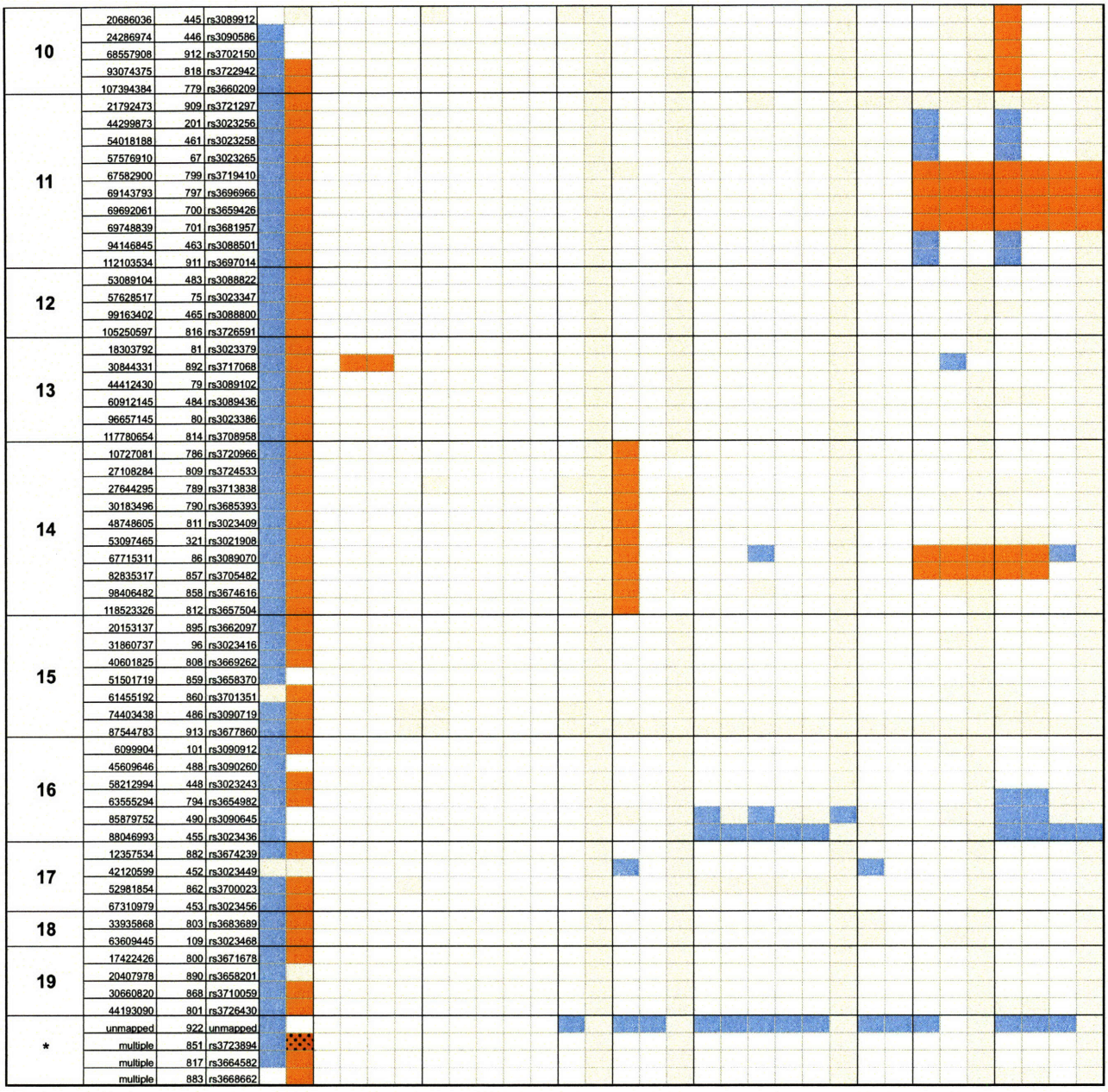


Table 4: Positively scored LOH markers from first-pass genomic screen

Listed are markers scored positively (AIF>3 or <0.3) in tumors but not control. The positive call rate was 1.1% (44/3828). In the second pass, 19 of the positive markers were individually genotyped by SBE on tumor and normal DNA. AIF was calculated using the paired data for each marker. The concordant positive call rate between the two passes was 37% (7/19).

Tumor	Assay ID	RefSNP ID	Chromosome	Nucleotide	AIF (1st pass)	AIF (2nd pass)	Concordant LOH call
800e	746	rs3657668	5	66812086	0.29	0.95	
	892	rs3717068	13	30844331	3.08		
800i	892	rs3717068	13	30844331	3.25		
803g	746	rs3657668	5	66812086	0.27		
818g	750	rs3709317	6	49704741	0.00		
840a	786	rs3720966	14	10727081	5.34	6.11	*
	809	rs3724533		27108284	5.21	1.24	*
	789	rs3713838		27644295	780.66	3275.75	*
	790	rs3685393		30183496	295.64	5.27	*
	811	rs3023409		48748605	6.37	3.20	*
	321	rs3021908		53097465	10.70	7.56	*
	86	rs3089070		67715311	5.18	5.15	*
	857	rs3705482		82835317	5.33	1.00	
	858	rs3674616		98406482	480.62	1.00	
	812	rs3657504		118523326	7.31	5.81	*
	452	rs3023449	17	42120599	0.26	0.60	
849b	4	rs3022839	1	137257629	3.09		
849c	86	rs3089070	14	67715311	0.00	3.90	
850d	750	rs3709317	6	49704741	0.00	4.66	
	879	rs3695889	9	86124796	0.02	1.00	
	452	rs3023449	17	42120599	0.24	0.75	
866h	850	rs3090435	6	146590775	3.30		
	201	rs3023256	11	44299873	0.21		
	461	rs3023258		54018188	0.00		
	67	rs3023265		57576910	0.00		
	463	rs3088501		94146845	0.00		
	911	rs3697014		112103534	0.19		
866i	830	rs3714631	2	151657817	668.48	0.67	
	833	rs3726717	3	52860548	98.83	1.00	
	892	rs3717068	13	30844331	0.19	1.06	
870a	766	rs3724779	8	117005581	23.23		
	768	rs3716232	9	16092806	180.31		
	769	rs3694785		16382724	169.88		
	770	rs4137954		98214467	4.87		
	445	rs3089912	10	20686036	77.82		
	446	rs3090586		24286974	89.13		
	912	rs3702150		68557908	3.70		
	818	rs3722942		93074375	7.43		
	779	rs3660209		107394384	5.20		
	201	rs3023256	11	44299873	0.25		
	461	rs3023258		54018188	0.01		
	67	rs3023265		57576910	0.20		
	463	rs3088501		94146845	0.01		
	911	rs3697014		112103534	0.25		
				Count:	**44**	**19**	**7**
				Total informative markers:	**3828**		

Loss of wild-type p53 on chromosome 11

Three of the analyzed tumors (866h, 866i, 870a) came from *Kras*^{LA2}*; p53*+/- mice. The p53 null allele in the F1 mice came from a p53+/- (B6) parent, generated from 129 ES cells in initial targeting and subsequent backcrossing to B6 for 20+ generations. Interestingly, in the normal tissue of these mice, the 4 markers (799, 797, 700, 701) within +0.34Mb and -1.82Mb around the *p53* locus on chromosome 11 were homozygous for the 129S4 allele (Figure 4). It appears that despite many generations of backcrossing to B6 mice, markers linked to the *p53* knock-out allele on chromosome 11 remain the 129S4 ancestry. Other SNPs on chromosome 11 were heterozygous in normal DNA of these *Kras*^{LA2}*; p53*+/- F1 mice. In the DNA from two tumors: 866h and 870a, the 129S4 alleles of these markers were lost, suggesting a selective loss of the whole chromosome 11 that contained wild-type *p53*. To test, PCR was performed to genotype the *p53* locus. As shown in Figure 5A, the ratio of the wild-type to mutant *p53* allele was decreased by about half in the PCR products from 866h and 870a, confirming the LOH data. The incomplete loss of wild-type signal could be due to a heterogeneous loss of *p53* among the tumor cells and/or contamination by normal stromal tissues in tumors.

Reduced level of wild-type Kras on chromosome 6

One sample, 866h, showed a loss of the B6 allele at marker rs3090435 (assay ID 850). on the distal arm of chromosome 6. Intriguingly, the marker is only +1.39Mb away from the *Kras* gene towards the telomere. We sought to examine if the B6 allele of *Kras*, which marked the wild-type copy of the gene, has also undergone concomitant allelic loss. In

creating the *Kras*^{LA2} allele, a novel *Hind*III restriction site was introduced in *Kras* exon 1 along with the point mutation. An SBE assay was designed to test for the new SNP at the engineered *Hind*III site (Figure 5B). As shown in Figure 5C, the wild-type allele (A) of *Kras* in 866h showed a reduction in relative signal when compared to tail control by ~2x, which was within LOH cutoff. The assay was also performed 849c, 840b, 850d, and 866i tumor DNA. Neither the SNP nor *Kras* exhibited a change in allelic ratios in these tumors (data not shown).

Comparison of LOH results to copy number data

Four lung tumor samples (840a, 870a, 849c, 866h) were chosen for copy number analysis using a ROMA platform. ROMA uses oligonucleotide microarrays to relatively quantify the low-complexity representations of tumor vs. normal genomes. Global chromosomal copy number changes were apparent in 840a, 870a, and 866h (Figure 6). These changes were summarized in Table 5 along with LOH data. Discrepancy was apparent: multiple chromosomes have reduced in copy number but no LOH was associated. These include chromosomes 4, 5, 11, and 15 in 840a, chromosome 19 in 870a, and chromosome 9 in 866h. Interestingly, chromosome 6 was increased in copy number in 866h (Figure 6), suggesting the changes in allelic ratio observed in *Kras* and rs3090435 reflect a gain of the mutant (129S4) copy instead of a loss of wild-type allele. The wild-type allele of *Kras* has been suggested to suppress tumor development (Zhang et al. 2001).

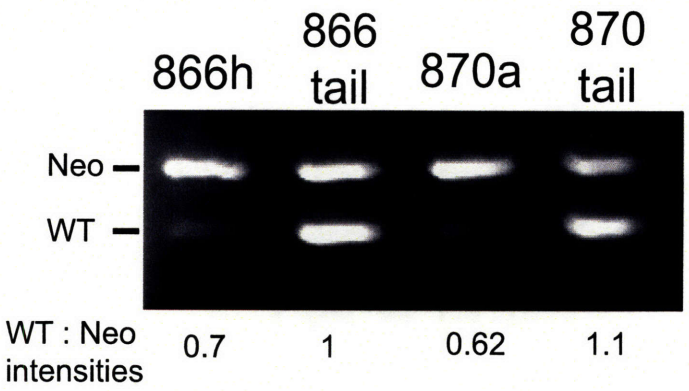
Figure 5: Genotyping results at *p53* and *Kras* loci

A) *p53* PCR genotyping was performed on tumor DNA from $Kras^{LA2}$; *p53*+/- mice to assess the relative intensities of wild-type *p53* to the knocked-out allele with an inserted Neo cassette.

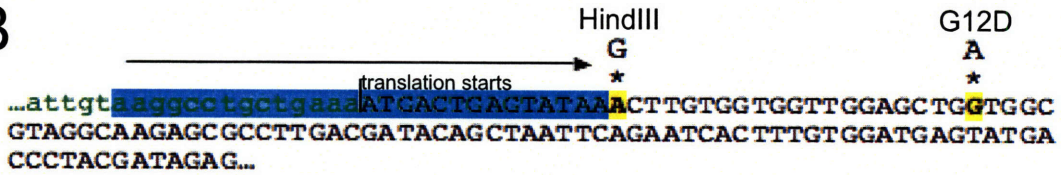
B) An SBE assay designed to assess the ratio of $Kras^{LA2}$ to wild-type alleles of *Kras*. The assay genotypes for the single-nucleotide difference on the $Kras^{LA2}$ allele at an engineered *Hind*III site closely linked to the G12D expressing point mutation. Sequence presented comes from the wild-type *Kras* gene. The two single-nucleotides changes introduced in making the $Kras^{LA2}$ allele are marked by asterisks. The genotyping SBE primer is highlighted in the sequence with strand direction marked by the arrow.

C) *Kras* genotyping results of 866h tumor DNA and control. Ratio of A/G peak heights reflects the wild-type to $Kras^{LA2}$ allelic ratio.

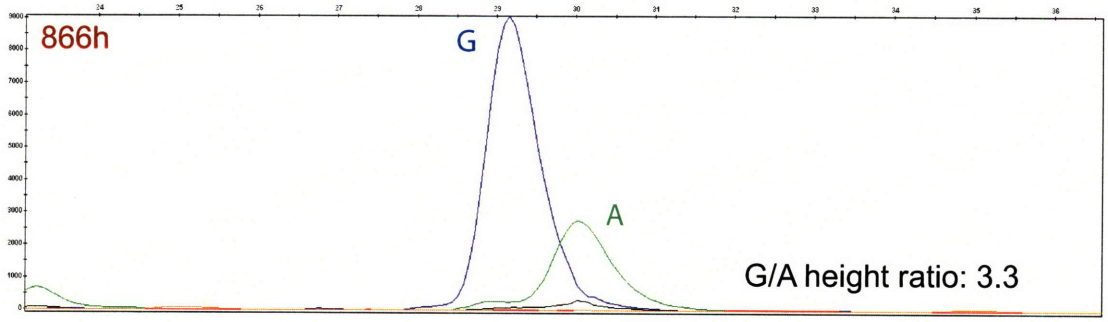
A


866h
866 tail
870a
870 tail
Neo
WT
WT : Neo intensities
0.7
1
0.62
1.1

B


HindIII
G
*
G12D
A
*
translation starts
...attgtaaggcctgctgaaaATGACTGACTATAAAACTTGTGGTGGTTGGAGCTGGTGGC
GTAGGCAAGAGCGCCTTGACGATACAGCTAATTCAGAATCACTTTGTGGATGAGTATGA
CCCTACGATAGAG...

C


866h
G
A
G/A height ratio: 3.3

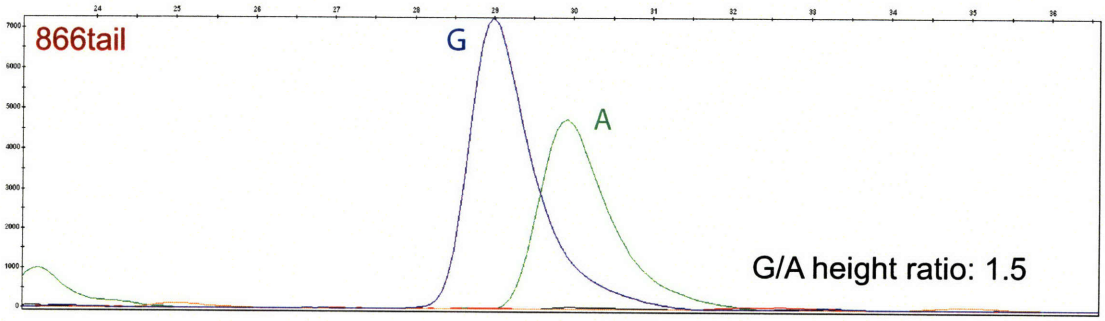

866tail
G
A
G/A height ratio: 1.5

Figure 6: Copy number analysis of tumor DNA by ROMA

Four genome-wide moving median plots showing fluorescence ratios of labeled tumor to labeled tail DNA. DNA from four tumors: 840a, 870a, 849c, and 866h were analyzed by ROMA. The Y-axis is the log10 fluoresence ratio and the X-axis is an index of the probes genomic order based on UCSC mouse May 2004 (mm5) annotated assembly. Data from each chromosome are labeled with same color. Chromosomes with deviated signals from either the positive or negative baselines are labeled.

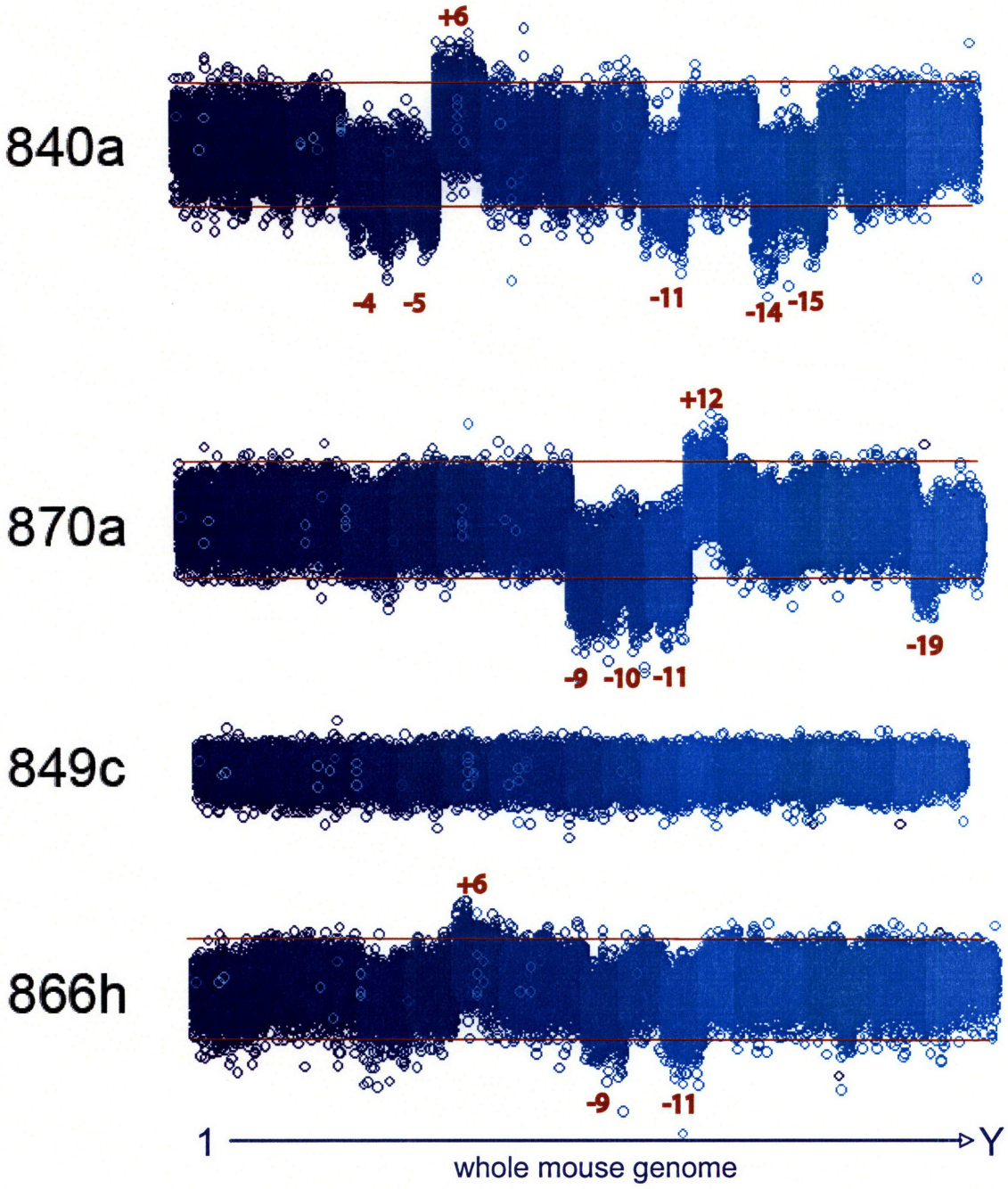

840a
+6
-4
-5
-11
-14
-15
870a
+12
-9
-10
-11
-19
849c
866h
+6
-9
-11
1
Y
whole mouse genome

Table 5: Comparison of chromosomal LOH and copy number changes as suggested by SNP genotyping and ROMA analysis

The same tumor DNA samples obtained from mice with indicated genotype were analyzed by SNPs and ROMA. The chromosomal LOH or copy number gains (+) and losses (-) observed respectively are listed.

Tumor ID	Mouse genotype	LOH	Copy number changes
840a	*Kras*^{LA2}	14	-4, -5, -11, -14, -15, +6
870a	*Kras*^{LA2}; *p53+/-*	9, 10, 11	-9, -10, -11, -19, +12
866h	*Kras*^{LA2}; *p53+/-*	11	-9, -11, +6

Discussion

Using SNPs as markers for genome-wide LOH screen in mouse cancer

This is the first report of a genome-wide LOH screen with SNPs in mouse tumor models. We described a protocol for performing genome-wide SNP genotyping by coupling SBE and capillary electrophoresis using SNaPshot™fluorescence chemistry (Applied Biosystems). The robustness and sensitivity of the method in simultaneously genotyping multiple markers have been demonstrated by various groups (Makridakis and Reichardt 2001; Norton et al. 2002; Ben-Avi et al. 2004). Implementation of the method for genome-wide genotyping has not been previously described. The current study extended the technique to a larger scale to analyze 147 SNPs throughout the mouse genome in 15 SBE reactions. The screening procedure positively identified LOH of chromosome 11 in two samples of mouse tumors, both of which were confirmed to involve a loss of the wild-type *p53* allele. Our data show SNP genotyping in F1 hybrid mice is a viable method for LOH screening. The limiting factor to perform genome-wide SNP screening in most labs is cost. Most other established methods for mouse genomic SNP genotyping involve specialized instruments, such as MALDI-TOF mass spectrometry(Wiltshire et al. 2003; Pletcher et al. 2004; Moran et al. 2006). On the other hand, the fixed cost of our screening method is essentially just the cost of primers, as only standard PCR and routine sequencing machines are needed. The presented protocol can likely be adoptable by most laboratories and can be further optimized to improve each SBE reaction to the highest accuracy level capable by same type of assay (Makridakis and Reichardt 2001; Norton et al. 2002). Sensitivity of the method needs to be tested on more markers in a dilution

experiment with heterozygous vs. different homozygous DNA. Directions for future improvements include further multiplexing of PCR, testing the assay utility with DNA amplified paraffin embedded samples, which would allow archived tissues to be characterized, and validating the use of the SNP panel on other mouse strains. Developing a bioinformatics approach to account for experimental noise may also improve the robustness of LOH assignments. Benchmarking the accuracy, sensitivity, and cost of this protocol against an established assay (Wiltshire et al. 2003; Pletcher et al. 2004; Moran et al. 2006) will provide an objective comparison of the different mouse SNP genotyping methods.

Implications on lung cancer genetics

Within the technical limit of our assay, the overall LOH rate was low in the experimental cohort of mouse lung tumors. The majority of the tumors analyzed maintained both parental alleles in all loci. Only 3/20 tumors exhibited various degree of LOH; all were histologically graded 3 or higher. Among the three tumors, LOH of chromosomes 9, 10, 11, and 14 was detected.

Combining LOH and CGH data

To compare LOH results to alterations in copy number, ROMA was used to characterize these 3 tumors along with one other sample, 849c. LOH may result through deletion, which is associated with a physical loss of the wild-type gene copy, or through non-disjunction-led chromosomal duplication, which might not result in a copy number change. In fact, most LOH events are not associated with copy number changes at least in humans (Huang et al. 2004; Beroukhim et al. 2006). In our case, limited by the

technicalities of the screening platforms, the reverse was observed: copy number losses of certain chromosomes were not associated with LOH (Table 5). Sample 840a provided the clearest example with ROMA showing reduced copy signals for chromosomes 4, 5, 11, 14, and 15 at nearly same magnitude, but only chromosome 14 showed LOH in our screen. There might be a a technical and a biological basis for the discrepancy. Technically, this discrepancy could be due to a difference in sensitivity of the two assays. While the quantitativeness of our assay was tested on two SNPs, other markers may behave differently. Furthermore, the sensitivity limit of ROMA is unclear. Although we were able to verify high-amplitude ROMA signal such as that from the *N-Myc* amplicon (see Chapter 2), subtle copy number differences might exist but not be detected by ROMA. The technical capability of the assay can be tested by confirming our LOH results using an independent assay, such as another SNP screen (Wade et al. 2002; Pletcher et al. 2004; Owens et al. 2005; Moran et al. 2006) or SSLP genotyping (Dietrich et al. 1994). We also attempted our screen on breast tumor DNA that have been previously characterized by array CGH (Chao et al. 2005). The concordance between known copy number losses and our LOH screening data appeared higher (Supplemental Figure 1), suggesting the discordance in the lung tumor samples may be real. Several biological reasons may explain the discordant copy number and allelotyping results. Since a tumor is believed to be a heterogeneous group of clonally selected cells, analysis using total DNA from a tumor is an assessment of the averaged genetic changes in its composite cells. As such, a chromosomal copy number reduction in total DNA implies a selection for the hemizygous state of that chromosome has occurred, while a lack of allelic loss suggests the choice of which chromosome to lose was random. This may

result if the chromosome involved contains one or more haplo-insufficient tumor suppressor genes, which can confer tumorigenicity when dosage is reduced through hemizygosity. Although tumor suppressor is classically thought to act through a 'two-hit' inactivation process, increasing number of haplo-insufficient tumor suppressor genes have been described. This model is based on the assumption that the tumor cells are diploid for the chromosomes without apparent changes by ROMA. However, if the cells are mostly tetraploid, a state that is believed to precede widespread aneuploidy (Fujiwara et al. 2005), which was in fact observed, the baseline for ROMA would no longer be two chromosomes. A decrease in ROMA signal could mean 4-1 or 4-2 chromosomes. Either one or both alleles could remain, leading to discordant results. While ROMA cannot distinguish ploidy, LOH analysis may lack the power to distinguish subtle allelic changes. Future experiments to complement karyotyping or FACS-based ploidy analysis on cells from the same tumors may resolve some issues.

Chromosome 11 in mouse lung tumorigenesis

Chromosome 11 LOH was seen in 2 tumors that came from $Kras^{LA2}$;*p53*+/- mice. In both cases, the lost chromosome contained the wild-type *p53* allele. This result is consistent to the observation by Jackson et al. that tumors induced by conditionally expressing $Kras^{G12D}$ and a *p53* deletion allele in the lung have also lost the wild-type copy of *p53*. In the present study, chromosome 11 LOH occurred in combination with LOH of chromosome 6 in one tumor and LOH of chromosomes 9 and 10 in the other. It was possible that loss of *p53* on chromosome 11 has provided a permissive environment for genome-wide aneuploidy. Consistent to the hypothesis, ROMA-generated copy number data showed widespread aneuploidy in these two tumors. *p53* has been implicated in

preventing tetraploid cells to proceed through the cell cycle, which frequently leads to aneuploidy (Lanni and Jacks 1998; Meraldi et al. 2002; Fujiwara et al. 2005). Interestingly, in tumor 840a from a $Kras^{LA2}$;*p53*+/+ mouse, chromosome 11 was among the chromosomes that showed copy number loss without LOH, which might imply haplo-insufficiency as discussed above. One candidate locus could again be *p53*. Although often lost biallelically, ~50% of mouse and human tumors with a *p53* mutation in fact retain the wild-type copy of the gene (Venkatachalam et al. 1998; Trkova et al. 2003). In addition, differences in *p53* dosage have been seen to affect tumor phenotype of a mouse model (Hemann et al. 2003). Finally, it is important to note that mouse chromosome 11 is very gene-rich, and its distal arm shares conserved synteny to the whole human chromosome 17, which is home to many disease-related genes included the tumor suppressors *Brca1* and *Nf1*. In lung cancer, 17p12-13 deletion is frequently observed (Balsara and Testa 2002). While it is tempting to relate our observations of chromosome 11 LOH and/or copy number loss to *p53* function, other loci may also be critical.

Note on genetic background

We generated our tumor samples using F1 mice from a 129xB6 cross. Strain-specific phenotypic differences appear to exist. $Kras^{LA}$ mice on a B6 background develop lung tumors at higher multiplicity than on 129S4 (Michel Dupage, personal communication). Molecularly, it is unclear whether tumors from C57BL/6J, 129S4/SvJae, and F1 mice have different types and rates of genetic changes. LOH resulting from mitotic recombination can get suppressed in F1 hybrids of different mouse strains (Shao et al. 2001). Furthermore, tumor susceptibility gene(s) specific to B6 and 129 strains have been mapped to chromosome 11 in a different tumor type (Reilly et al. 2004). Identity and

function of the modifiers are unclear but may affect LOH frequency of chromosome 11 in tumors of other types as well.

Concluding note

Using a multiplexed SNaPshot™genotyping protocol for genome-wide SNP detection, we have detected LOH along several chromosomes on mouse tumors. Further optimization will improve both the throughput and accuracy of the assay. Aside from LOH identification in tumors, a protocol for SNP genotyping method for mouse can also be useful for other purposes such as positional cloning of modifiers in cancer and in other diseases.

Materials and Methods

Tumor DNA isolation

KRas $^{LA2/+}$ mice on a 129S4/SvJae background were crossed to wild-type, *p53*+/- or *p53 R270H/+* mice on C57BL6J background to obtain F1 progeny. Lung tumors were dissected from the lungs of F1 mice between 5-8 months of age. One portion of each tumor was fixed in formalin, sectioned in paraffin, and stained in hematoxylin and eosin. Histological grading of each tumor was assigned based on a 1-5 scale as described previously (Jackson et al. 2005). Remaining tumor material was stored at -80°C prior to DNA isolation. DNA was extracted from thawed tissues using reagents and protocols in Puregene DNA isolation kit (Gentra Systems, Inc.).

SNP genotyping

SNP genotype was assessed using a single-base extension (SBE) method (Sokolov 1990). SNP targets were first amplified by PCR using 2ng genomic DNA, 0.3mM dNTPs, 0.4mM PCR primers, 1x GeneAmp Gold buffer and 0.5U of AmpliTaq Gold polymerase (Applied Biosystems) under the following conditions: 95°C for 9mins, 35 cycles of 94°C for 30s, 55°C for 30s and 72°C for 45s, and a final extension for 5 min at 72°C. Unincorporated PCR primers and dNTPs were removed with 2U exonuclease I (Applied Biosystems) and 2U shrimp alkaline phosphatase (Applied Biosystems) at 37°C for 1 hour. 0.15-1.5pmol of pmol of SBE primer was added to the treated PCR templates with SNaPshot Multiplex Ready Reaction Mix (Applied Biosystems) and cycled 25 times at 96°C for 10s, 50°C for 5s and 60°C for 30s. Post-extension products were treated with 0.25U of shrimp alkaline phosphatase (Applied Biosystems) to remove unincorporated ddNTPs. Final products were mixed with 0.25ml of Liz-120 size standards (Applied Biosystems) and ran on the Applied Biosystems 3700 DNA Analyzer.

To test markers for polymorphism in 129S4/SvJae vs. C57BL6J strains, SNPs were identified in public databases generated from two large-scale SNP identification efforts (Germer et al. 2000; Lindblad-Toh et al. 2000b) and tested individually using primers listed in Supplementary Table 1. To perform genome-wide analysis, 147 validated SNPs were chosen. PCR were performed in singlet or duplex on 96-well plates and pooled according to their SBE group as in Table 2. Pooled products were purified and concentrated using multi-well PCR purification kit (Qiagen) for SBE reactions as described above.

Positional information of each SNP is obtained through sequence blat search of the UCSC Mouse Feb 2006 (mm8) assembly.

SNP Data Analysis

First round data analysis was done with Applied Biosystems GeneMapper 3.7 software. By comparing electrographs of multiplex and individual SBE reactions, peaks corresponding to each SNP in a multiplex lane were manually identified. Marker boundaries were set as bins to guide the software to automatically assign genotypes for peaks that fall within. All computer-generated assignments were inspected manually before final analysis.

The height ratio of the peaks associated with the two alleles of each SNP was calculated. Allelic imbalance factor (AIF) of each marker in a sample is determined as follows:

$AIF = (H_1/H_2)/(H_{1ref}/H_{2ref})$,

where H_1 = Sample peak height associated with 129S4 allele,

H_2 = Sample peak height associated with C57BL6J allele,

H_{1ref} = Reference peak height associated with 129S4 allele,

and H_{2ref} = Reference peak height associated with C57BL6J allele.

Reference peak heights were calculated by averaging data from three independent normal F1 DNA controls. A positive call for LOH was made when AIF is >3 or <0.3

LOH assessment of p53 and Kras loci

Ratio of wild-type to knock-out (with neomycin cassette) allele of p53 in tumors was assessed using the standard tail genotyping protocol with the following primers: p53x6.5: ACAGCGTGGTGGTACCTTAT, p53x7: TATACTCAGAGCCGGCCT, and Neo18.5: TCCTCGTGCTTTACGGTATC, yielding 375bp and 525bp products corresponding to the wild-type and knocked-out alleles. Number of PCR cycles was reduced to 22 in genotyping tumors.

To assess the ratio of Kras LA2 to wild-type alleles, an SBE assay was used to genotype the single-nucleotide difference in the LA2 allele at the novel HindIII site near the G12D expressing point mutation. The following PCR primers were used: LA-F1: GGTTACTCTGTACATCTGTAGTCACTG and LA-R1: AAGCGCACGCAGACTGTAG. SBE primer sequence was AAGGCCTGCTGAAAATGACTGAGTATAA.

DNA copy number analysis

ROMA was employed to assess genome-wide copy number changes and was performed as described in Chapter Two of this thesis.

Acknowledgements

The authors would like to thank Kevin Shannon for generous gifts of breast tumor and cell line DNA samples used in experiment described in the supplemental section. We also thank Charlie Whittaker and Ray Farhner for computing support, Denise Crowley and

Mike Brown for histology preparation, and Rob Bronson for pathology consultation. T.J. is a Howard Hughes Investigator.

References

Balsara, B.R. and Testa, J.R. 2002. Chromosomal imbalances in human lung cancer. *Oncogene* **21**(45): 6877-6883.

Ben-Avi, L., Durst, R., Shpitzen, S., Leitersdorf, E., and Meiner, V. 2004. Apolipoprotein E genotyping: accurate, simple, high throughput method using ABI Prism SNaPshot Multiplex System. *J Alzheimers Dis* **6**(5): 497-501.

Benavides, F., Conti, C.J., LaCava, M., Flores, M., Glasscock, E., Sternik, G., Gimenez-Conti, I.B., Johnston, D.A., Dunsford, H.A., Goldstein, L.S., and Rodriguez, L.V. 2003. Loss of heterozygosity analysis of mouse pulmonary adenomas induced by coal tar. *Environ Mol Mutagen* **41**(4): 300-308.

Beroukhim, R., Lin, M., Park, Y., Hao, K., Zhao, X., Garraway, L.A., Fox, E.A., Hochberg, E.P., Mellinghoff, I.K., Hofer, M.D., Descazeaud, A., Rubin, M.A., Meyerson, M., Wong, W.H., Sellers, W.R., and Li, C. 2006. Inferring loss-of-heterozygosity from unpaired tumors using high-density oligonucleotide SNP arrays. *PLoS Comput Biol* **2**(5): e41.

Chao, R.C., Pyzel, U., Fridlyand, J., Kuo, Y.M., Teel, L., Haaga, J., Borowsky, A., Horvai, A., Kogan, S.C., Bonifas, J., Huey, B., Jacks, T.E., Albertson, D.G., and Shannon, K.M. 2005. Therapy-induced malignant neoplasms in Nf1 mutant mice. *Cancer Cell* **8**(4): 337-348.

Dietrich, W.F., Radany, E.H., Smith, J.S., Bishop, J.M., Hanahan, D., and Lander, E.S. 1994. Genome-wide search for loss of heterozygosity in transgenic mouse tumors reveals candidate tumor suppressor genes on chromosomes 9 and 16. *Proc Natl Acad Sci U S A* **91**(20): 9451-9455.

Frazer, K.A., Wade, C.M., Hinds, D.A., Patil, N., Cox, D.R., and Daly, M.J. 2004. Segmental phylogenetic relationships of inbred mouse strains revealed by fine-scale analysis of sequence variation across 4.6 mb of mouse genome. *Genome Res* **14**(8): 1493-1500.

Fujiwara, T., Bandi, M., Nitta, M., Ivanova, E.V., Bronson, R.T., and Pellman, D. 2005. Cytokinesis failure generating tetraploids promotes tumorigenesis in p53-null cells. *Nature* **437**(7061): 1043-1047.

Germer, S., Holland, M.J., and Higuchi, R. 2000. High-throughput SNP allele-frequency determination in pooled DNA samples by kinetic PCR. *Genome Res* **10**(2): 258-266.

Grupe, A., Germer, S., Usuka, J., Aud, D., Belknap, J.K., Klein, R.F., Ahluwalia, M.K., Higuchi, R., and Peltz, G. 2001. In silico mapping of complex disease-related traits in mice. *Science* **292**(5523): 1915-1918.

Hemann, M.T., Fridman, J.S., Zilfou, J.T., Hernando, E., Paddison, P.J., Cordon-Cardo, C., Hannon, G.J., and Lowe, S.W. 2003. An epi-allelic series of p53 hypomorphs created by stable RNAi produces distinct tumor phenotypes in vivo. *Nat Genet* **33**(3): 396-400.

Herzog, C.R., Devereux, T.R., Pittman, B., and You, M. 2002. Carcinogenic induction directs the selection of allelic losses in mouse lung tumorigenesis. *Cancer Res* **62**(22): 6424-6429.

Huang, J., Wei, W., Zhang, J., Liu, G., Bignell, G.R., Stratton, M.R., Futreal, P.A., Wooster, R., Jones, K.W., and Shapero, M.H. 2004. Whole genome DNA copy number changes identified by high density oligonucleotide arrays. *Hum Genomics* **1**(4): 287-299.

Jackson, E.L., Olive, K.P., Tuveson, D.A., Bronson, R., Crowley, D., Brown, M., and Jacks, T. 2005. The differential effects of mutant p53 alleles on advanced murine lung cancer. *Cancer Res* **65**(22): 10280-10288.

Johnson, L., Mercer, K., Greenbaum, D., Bronson, R.T., Crowley, D., Tuveson, D.A., and Jacks, T. 2001. Somatic activation of the K-ras oncogene causes early onset lung cancer in mice. *Nature* **410**(6832): 1111-1116.

Knudson, A.G., Jr. 1971. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* **68**(4): 820-823.

Lanni, J.S. and Jacks, T. 1998. Characterization of the p53-dependent postmitotic checkpoint following spindle disruption. *Mol Cell Biol* **18**(2): 1055-1064.

Lindblad-Toh, K., Winchester, E., Daly, M.J., Wang, D.G., Hirschhorn, J.N., Laviolette, J.P., Ardlie, K., Reich, D.E., Robinson, E., Sklar, P., Shah, N., Thomas, D., Fan, J.B., Gingeras, T., Warrington, J., Patil, N., Hudson, T.J., and Lander, E.S. 2000. Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nat Genet* **24**(4): 381-386.

Makridakis, N.M. and Reichardt, J.K. 2001. Multiplex automated primer extension analysis: simultaneous genotyping of several polymorphisms. *Biotechniques* **31**(6): 1374-1380.

Meraldi, P., Honda, R., and Nigg, E.A. 2002. Aurora-A overexpression reveals tetraploidization as a major route to centrosome amplification in p53-/- cells. *Embo J* **21**(4): 483-492.

Moran, J.L., Bolton, A.D., Tran, P.V., Brown, A., Dwyer, N.D., Manning, D.K., Bjork, B.C., Li, C., Montgomery, K., Siepka, S.M., Vitaterna, M.H., Takahashi, J.S., Wiltshire, T., Kwiatkowski, D.J., Kucherlapati, R., and Beier, D.R. 2006. Utilization of a whole genome SNP panel for efficient genetic mapping in the mouse. *Genome Res* **16**(3): 436-440.

Norton, N., Williams, N.M., Williams, H.J., Spurlock, G., Kirov, G., Morris, D.W., Hoogendoorn, B., Owen, M.J., and O'Donovan, M.C. 2002. Universal, robust, highly quantitative SNP allele frequency measurement in DNA pools. *Hum Genet* **110**(5): 471-478.

Owens, S.E., Broman, K.W., Wiltshire, T., Elmore, J.B., Bradley, K.M., Smith, J.R., and Southard-Smith, E.M. 2005. Genome-wide linkage identifies novel modifier loci of aganglionosis in the Sox10Dom model of Hirschsprung disease. *Hum Mol Genet* **14**(11): 1549-1558.

Petkov, P.M., Ding, Y., Cassell, M.A., Zhang, W., Wagner, G., Sargent, E.E., Asquith, S., Crew, V., Johnson, K.A., Robinson, P., Scott, V.E., and Wiles, M.V. 2004. An efficient SNP system for mouse genome scanning and elucidating strain relationships. *Genome Res* **14**(9): 1806-1811.

Pletcher, M.T., McClurg, P., Batalov, S., Su, A.I., Barnes, S.W., Lagler, E., Korstanje, R., Wang, X., Nusskern, D., Bogue, M.A., Mural, R.J., Paigen, B., and Wiltshire, T. 2004. Use of a dense single nucleotide polymorphism map for in silico mapping in the mouse. *PLoS Biol* **2**(12): e393.

Radany, E.H., Hong, K., Kesharvarzi, S., Lander, E.S., and Bishop, J.M. 1997. Mouse mammary tumor virus/v-Ha-ras transgene-induced mammary tumors exhibit strain-specific allelic loss on mouse chromosome 4. *Proc Natl Acad Sci U S A* **94**(16): 8664-8669.

Reilly, K.M., Tuskan, R.G., Christy, E., Loisel, D.A., Ledger, J., Bronson, R.T., Smith, C.D., Tsang, S., Munroe, D.J., and Jacks, T. 2004. Susceptibility to astrocytoma in mice mutant for Nf1 and Trp53 is linked to chromosome 11 and subject to epigenetic effects. *Proc Natl Acad Sci U S A* **101**(35): 13008-13013.

Shao, C., Stambrook, P.J., and Tischfield, J.A. 2001. Mitotic recombination is suppressed by chromosomal divergence in hybrids of distantly related mouse strains. *Nat Genet* **28**(2): 169-172.

Sokolov, B.P. 1990. Primer extension technique for the detection of single nucleotide in genomic DNA. *Nucleic Acids Res* **18**(12): 3671.

Sweet-Cordero, A., Mukherjee, S., Subramanian, A., You, H., Roix, J.J., Ladd-Acosta, C., Mesirov, J., Golub, T.R., and Jacks, T. 2005. An oncogenic KRAS2 expression signature identified by cross-species gene-expression analysis. *Nat Genet* **37**(1): 48-55.

Syvanen, A.C. 2001. Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nat Rev Genet* **2**(12): 930-942.

Trkova, M., Foretova, L., Kodet, R., Hedvicakova, P., and Sedlacek, Z. 2003. A Li-Fraumeni syndrome family with retained heterozygosity for a germline TP53 mutation in two tumors. *Cancer Genet Cytogenet* **145**(1): 60-64.

Venkatachalam, S., Shi, Y.P., Jones, S.N., Vogel, H., Bradley, A., Pinkel, D., and Donehower, L.A. 1998. Retention of wild-type p53 in tumors from p53 heterozygous mice: reduction of p53 dosage can promote cancer formation. *Embo J* **17**(16): 4657-4667.

Wade, C.M., Kulbokas, E.J., 3rd, 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**(6915): 574-578.

Waterston, R.H. Lindblad-Toh, K. Birney, E. Rogers, J. Abril, J.F. Agarwal, P. Agarwala, R. Ainscough, R. Alexandersson, M. An, P. Antonarakis, S.E.

Attwood, J. Baertsch, R. Bailey, J. Barlow, K. Beck, S. Berry, E. Birren, B. Bloom, T. Bork, P. Botcherby, M. Bray, N. Brent, M.R. Brown, D.G. Brown, S.D. Bult, C. Burton, J. Butler, J. Campbell, R.D. Carninci, P. Cawley, S. Chiaromonte, F. Chinwalla, A.T. Church, D.M. Clamp, M. Clee, C. Collins, F.S. Cook, L.L. Copley, R.R. Coulson, A. Couronne, O. Cuff, J. Curwen, V. Cutts, T. Daly, M. David, R. Davies, J. Delehaunty, K.D. Deri, J. Dermitzakis, E.T. Dewey, C. Dickens, N.J. Diekhans, M. Dodge, S. Dubchak, I. Dunn, D.M. Eddy, S.R. Elnitski, L. Emes, R.D. Eswara, P. Eyras, E. Felsenfeld, A. Fewell, G.A. Flicek, P. Foley, K. Frankel, W.N. Fulton, L.A. Fulton, R.S. Furey, T.S. Gage, D. Gibbs, R.A. Glusman, G. Gnerre, S. Goldman, N. Goodstadt, L. Grafham, D. Graves, T.A. Green, E.D. Gregory, S. Guigo, R. Guyer, M. Hardison, R.C. Haussler, D. Hayashizaki, Y. Hillier, L.W. Hinrichs, A. Hlavina, W. Holzer, T. Hsu, F. Hua, A. Hubbard, T. Hunt, A. Jackson, I. Jaffe, D.B. Johnson, L.S. Jones, M. Jones, T.A. Joy, A. Kamal, M. Karlsson, E.K. Karolchik, D. Kasprzyk, A. Kawai, J. Keibler, E. Kells, C. Kent, W.J. Kirby, A. Kolbe, D.L. Korf, I. Kucherlapati, R.S. Kulbokas, E.J. Kulp, D. Landers, T. Leger, J.P. Leonard, S. Letunic, I. Levine, R. Li, J. Li, M. Lloyd, C. Lucas, S. Ma, B. Maglott, D.R. Mardis, E.R. Matthews, L. Mauceli, E. Mayer, J.H. McCarthy, M. McCombie, W.R. McLaren, S. McLay, K. McPherson, J.D. Meldrim, J. Meredith, B. Mesirov, J.P. Miller, W. Miner, T.L. Mongin, E. Montgomery, K.T. Morgan, M. Mott, R. Mullikin, J.C. Muzny, D.M. Nash, W.E. Nelson, J.O. Nhan, M.N. Nicol, R. Ning, Z. Nusbaum, C. O'Connor, M.J. Okazaki, Y. Oliver, K. Overton-Larty, E. Pachter, L. Parra, G. Pepin, K.H. Peterson, J. Pevzner, P. Plumb, R. Pohl, C.S. Poliakov, A. Ponce, T.C. Ponting, C.P. Potter, S. Quail, M. Reymond, A. Roe, B.A. Roskin, K.M. Rubin, E.M. Rust, A.G. Santos, R. Sapojnikov, V. Schultz, B. Schultz, J. Schwartz, M.S. Schwartz, S. Scott, C. Seaman, S. Searle, S. Sharpe, T. Sheridan, A. Shownkeen, R. Sims, S. Singer, J.B. Slater, G. Smit, A. Smith, D.R. Spencer, B. Stabenau, A. Stange-Thomann, N. Sugnet, C. Suyama, M. Tesler, G. Thompson, J. Torrents, D. Trevaskis, E. Tromp, J. Ucla, C. Ureta-Vidal, A. Vinson, J.P. Von Niederhausern, A.C. Wade, C.M. Wall, M. Weber, R.J. Weiss, R.B. Wendl, M.C. West, A.P. Wetterstrand, K. Wheeler, R. Whelan, S. Wierzbowski, J. Willey, D. Williams, S. Wilson, R.K. Winter, E. Worley, K.C. Wyman, D. Yang, S. Yang, S.P. Zdobnov, E.M. Zody, M.C. and Lander, E.S. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**(6915): 520-562.

Wiltshire, T., Pletcher, M.T., Batalov, S., Barnes, S.W., Tarantino, L.M., Cooke, M.P., Wu, H., Smylie, K., Santrosyan, A., Copeland, N.G., Jenkins, N.A., Kalush, F., Mural, R.J., Glynne, R.J., Kay, S.A., Adams, M.D., and Fletcher, C.F. 2003. Genome-wide single-nucleotide polymorphism analysis defines haplotype patterns in mouse. *Proc Natl Acad Sci U S A* **100**(6): 3380-3385.

Wu, Y., Renard, C.A., Apiou, F., Huerre, M., Tiollais, P., Dutrillaux, B., and Buendia, M.A. 2002. Recurrent allelic deletions at mouse chromosomes 4 and 14 in Myc-induced liver tumors. *Oncogene* **21**(10): 1518-1526.

Zhang, J., Hunter, K.W., Gandolph, M., Rowe, W.L., Finney, R.P., Kelley, J.M., Edmonson, M., and Buetow, K.H. 2005. A high-resolution multistrain haplotype analysis of laboratory mouse genome reveals three distinctive genetic variation patterns. *Genome Res* **15**(2): 241-249.

Zhang, Z., Wang, Y., Vikis, H.G., Johnson, L., Liu, G., Li, J., Anderson, M.W., Sills, R.C., Hong, H.L., Devereux, T.R., Jacks, T., Guan, K.L., and You, M. 2001. Wildtype Kras2 can inhibit lung carcinogenesis in mice. *Nat Genet* **29**(1): 25-33.

Supplemental Information

Supplemental Figure 1: Genome-wide LOH screening of mouse breast tumors and tumor cell line

Grid illustrating data from the SNP screen columns: columns represent different samples; rows are markers organized by the chromosomal positions from UCSC Feb 2006 (mm8) assembly. AIF scores are labeled in each box. n/a=non-informative markers. Markers with positive AIF score are colored as follows: orange=AIF>3 (loss of B6 allele), blue=AIF<0.3 (loss of 129S4 allele). Tumor DNA came from breast tumors (B1, B2) or breast tumor cell lines (B1CL and C2CL) from Nf1+/- mice treated with radiation and cyclophosphamide (Chao et al. 2005). DNA was gift from Kevin Shannon. B1 DNA expressed copy number losses in whole or large regions of chromosomes 8, 11, 12, 14, 19 when profiled by array CGH. Most loci on these chromosomes scored positively for LOH.

Chromosome	Nucleotide	SNP ID	Sample ID			
			B1	B1CL	B2	B2CL
1	16141960	713	1.17			0.9
	33560855	714	0.906	0.88	0.99	0.94
	34060397	715	0.663			0.95
	53107439	898	1.159	1.29	1.59	1.29
	66668249	600	1.351			
	97464195	821	1.019			
	111247157	822			0.95	0.93
	121777951	721	1.325		1.35	1.18
	138360597	4	1.066	1.19	1.22	1.22
	155719123	823	1.066	0.91	1.06	0.98
	177919260	900	1.33	1.05	1.27	1.07
2	12070016	722	0.883		0.88	1.07
	12070228	723	1.03		1.01	1.08
	28621371	724	0.938		0.94	1.13
	45954896	726			1.57	1.13
	46472711	727	0.82			0.57
	69373291	405	1.215	0.89	1.18	0.82
	107222011	826	1.011			0.92
	118395454	827	1.4		1.14	1.2
	152709331	830	0.823		0.86	0.87
	168774394	730	1.102	1	1.21	1.28
	169277255	731			1.25	1.33
3	12531240	732	0.698			0.85
	28779003	831	0.95	1.18	1	0.89
	43382257	832			1.21	1.16
	54179064	833	0.658			1.25
	66900560	20	1.42	0.93	1.09	1
	80194307	834	1.259			0.87
	96226609	835	1.069		0.95	0.98
	134326444	412	0.878	0.83	0.91	0.85
	144714634	413	1.457	1.75	0.81	0.9
4	8073046	471	1.122		0.65	1.01
	26596737	738	0.94		1.54	
	53874480	839	1.471			1.25
	61518885	840	1.312			1
	74855800	415	1.109			0.88
	85139822	841	1.507			
	99041841	478	1.054	0.72	0.85	0.87
	130811701	740	1.427	1.27	1.37	1.24
	131376085	741	0.832	0.85	0.86	0.92
	140240863	218	0.984	1.01	0.74	0.84
5	22413727	32	0.885			1.08
	23730340	742	1.164	1.38	1.11	0.99
	49773577	422	1.069			1.53
	50105985	745	0.873		0.71	0.84
	65816926	746				0.44
	79989864	843	1.239	1.39	1.23	0.79
	99985513	844				
	116250135	37	1.053	0.83	0.91	0.75
	124509140	33	1.326		1	1.06
	131433208	748	0.899	0.91	1.04	0.98
6	18655079	426	1.06	0.94	1.26	1.04
	35969149	845	1.119		1.05	1.16
	50227000	750	0.941	0.79	0.95	0.96
	133171923	922	1.109	0.98	0.96	0.94
	148155268	850	0.934		1.17	0.97
7	4961	820				
	28228061	606				0.77
	43598764	851	0.748			0.83
	78082981	854	1.09	1.06	1.27	1.37
	88768207	435	0.85		1.09	1.22
	108372753	436	0.956			0.85
	117242899	756	1.392		1.05	1.09
	117499219	757	0.759	0.94	0.93	1.02
	121080866	339	0.656	0.67	0.82	0.85

8	12714613	118	0.483		1.01	1.1
	25189627	761		1.06	1.32	1.23
	33689228	439				
	56995287	762		0.65	0.95	0.99
	57031272	763		1.28	1.21	0.97
	89521985	51			1.04	1.2
	115796380	766			0.85	0.93
	116457023	767			0.88	0.99
9	16547432	768	1.299	1.2	1.38	1.23
	16834053	769	0.951		0.83	0.83
	89080475	879	1.195		1.1	1.27
	100802904	770	1.004	1.37	1.02	1.06
10	21831681	445	0.96			0.92
	25479871	446	1.06	0.66	0.76	0.98
	58804746	448	0.961		0.97	0.88
	71047972	912	1.188		0.96	1.13
	96023689	818	1.241	1.28	1.18	1
	110493528	779	0.954	1.06	0.96	0.96
11	21791444	909	0.668	0.75	0.91	
	44081495	201		1.08	1.05	1.15
	53799952	461				0.99
	57358532	67			0.87	0.71
	67338072	799				
	68899165	797			0.99	0.94
	69447433	700			1.72	
	69504211	701			1.31	1.31
	93914069	463	1.257	1.18	0.99	0.91
	111951607	911			1.04	0.98
12	47688508	483		0.9	1.04	0.81
	52059690	75	0.455	1.45	2.08	1.52
	94317217	465			1.53	1.19
	100385178	816			0.76	0.7
	114701971	817		0.95	1.16	1.03
13	18116231	81			0.95	0.96
	30724973	892	0.442		2.93	2.57
	44450246	79			1.21	0.95
	60683731	484		1.19	0.78	1.05
	93814910	80			2.29	2.07
	115468718	814	0.441	0.89	0.54	0.65
14	7269139	786	0.843	1.36	1.06	0.84
	22818444	809	1.165			
	23356191	789	0.964	0.83		
	25893857	790				1.12
	41573852	811	1.183	2.34	1.4	1.29
	45227636	321	0.868			0.91
	60528292	86	1.022	1.38	0.49	1.03
	75859013	857	0.975		1.02	1.17
	92333517	858	0.724			0.89
	112509253	812				0.98
15	20402168	895	0.839			1.24
	32440466	96	1.308			
	41170909	808	1.109		1.6	1.2
	52494598	859	1.024		0.92	1.14
	62885156	860	1.083		0.78	0.89
	76278665	486	0.876	0.99	1.41	1.08
	89405702	913				
16	5677800	101	0.699		0.69	0.87
	46036961	488	0.993	1.97	0.98	1.15
	64302751	794	1.145	0.89	1.04	1.01
	87237069	490	0.974	1.02	1.12	1.11
	89495624	455	1.453	1.04	0.52	0.93
17	12294400	882	1.295	1.32	1.13	1.28
	20829561	883	1.205	0.98	0.92	0.73
	42022515	452	1.037		0.9	
	53137039	862			0.85	1
	67644893	453			1.14	0.89
18	33996227	803	1.128	1.87	0.99	1.03
	63588982	109	1.412			0.94
19	3777	800	1.197	0.79	0.99	1.23
	20523042	890	1.534			1.25
	30292072	868		0.73	0.66	0.64
	44235554	801		0.97	1.31	1.07

Supplemental Table 1: Validation of 358 SNP markers in 129S4/SvJae vs. C57BL6J mouse strains

Genotyping results of 358 SNPs from public databases were obtained using a single-base extension method. Data were obtained from genotyping DNA from 129S4/SvJae, C57BL6J and 129S4/SvJae x C57BL6J F1 mice. The 147 highlighted SNPs that distinguish between the strains were chosen for the genome-wide screen. Chromosomal coordinates are based on UCSC mouse Feb 2006 (mm8) annotated assembly. Primer sequences used for PCR and SBE are listed in Supplemental Table 1.

Assay ID	RedSNP ID	Chromosome	Nucleotide	129S4	B6	F1(129S4xB6)
712	rs3691476	1	15866460	A	CA	CA
713	rs3684370	1	15963812	T	C	CT
714	rs3662163	1	32822284	T	C	CT
715	rs3704392	1	33320247	T	G	GT
400	rs3090110	1	45907197	C	T	-
898	rs3667466	1	51794155	C	T	CT
600	rs3022802	1	64979783	G	T	GT
899	rs3694327	1	79643870	C	C	CA
716	rs3716254	1	93320412	T	A	A
717	rs3657283	1	93320420	T	C	C
821	rs3704980	1	97461966	T	G	GT
822	rs3685188	1	110654171	T	C	CT
718	rs3713835	1	111883427	C	T	T
719	rs3662850	1	111887865	C	T	T
721	rs3664018	1	121061603	C	T	CT
720	rs3689749	1	121090273	G	A	A
603	rs3022832	1	127102172	C	-	G
601	rs3090765	1	127691908	T	G	G
4	rs3022839	1	137257629	G	A	GA
823	rs3713224	1	154052109	A	T	AT
402	rs3022851	1	171617363	T	C	-
403	rs3022871	1	174794383	C	C	-
900	rs3716435	1	175899264	C	T	CT
824	rs3704926	1	184806157	CT	T	CT
722	rs3681847	2	11984102	T	C	CT
723	rs4137557	2	11984314	T	C	CT
724	rs3664805	2	28174259	C	T	CT
10	rs3022883	2	37572959	G	G	G
726	rs3667376	2	45472806	C	T	CT
727	rs3681675	2	45990821	T	C	CT
404	rs3089489	2	51552294	A	A	-
405	rs3022887	2	68883673	G	A	GA
901	rs4137272	2	79034808	G	G	GA
406	rs3022888	2	84403279	A	A	-
825	rs3669855	2	93586287	G	G	GA
902	rs3670817	2	94071001	C	C	CT
728	rs3665286	2	94807655	G	G	G
729	rs3692288	2	95029801	C	C	T
826	rs3663534	2	106743641	T	C	CT
827	rs3090608	2	117816539	C	T	CT
18	rs3022895	2	119273525	G	GA	GA
828	rs3706262	2	119748062	A	G	GA
829	rs3660910	2	133743791	T	C	CT
408	rs3022910	2	144973062	C	T	-
830	rs3714631	2	151657817	C	T	CT
409	rs3089031	2	155762997	C	-	-
410	rs3022939	2	159674008	C	-	-
730	rs3666331	2	167851555	A	G	GA
731	rs3674631	2	168354454	T	C	CT
732	rs3706063	3	12332500	T	C	CT
733	rs3672323	3	12812319	A	C	CA
831	rs3676476	3	28110853	T	C	CT
411	rs3022953	3	37599994	C	-	-
832	rs3669022	3	42396732	C	T	CT
833	rs3726717	3	52860548	A	G	GA
20	rs3022960	3	65418042	G	A	GA
834	rs3689513	3	78626985	C	T	CT
835	rs3669413	3	93426396	C	A	CA
836	rs3722681	3	109900215	T	CT	-
903	rs3672398	3	112220863	T	AT	AT
212	rs3022965	3	114011713	G	-	-
904	rs3715748	3	119894667	A	GA	GA
837	rs3691246	3	122466065	G	GA	-

734	rs3714750	3	126334813	T	A	A
412	rs3089257	3	130719076	T	C	CT
413	rs3090381	3	140722145	A	G	GA
736	rs3656469	3	143498147	CT	C	CT
737	rs3706436	3	143786333	GT	TA	GTA
471	rs3022975	4	8073146	C	T	CT
414	rs3091112	4	8073150	GA	GA	-
738	rs3711350	4	26559488	A	G	GA
739	rs3684156	4	27375172	T	C	CTA
838	rs3665192	4	39271744	T	G	-
213	rs3088670	4	47484118	T	C	CT
839	rs3667625	4	54215713	A	G	GA
840	rs3694308	4	62440427	C	A	CA
415	rs3022979	4	75874108	C	A	CA
841	rs3707288	4	85885441	C	A	CA
709	rs3713394	4	86423755	T	T	CT
707	rs4135993	4	88399014	G	G	GA
708	rs3714181	4	88415879	T	T	CT
705	rs3680265	4	88774811	C	C	G
706	rs3696308	4	88775106	T	T	T
711	rs3686204	4	89249504	C	C	CT
710	rs3659287	4	89742501	C	C	CT
416	rs3090804	4	92499971	C	T	-
478	rs3022989	4	99679747	G	A	GA
417	rs3088455	4	99679889	C	C	-
842	rs3658845	4	115459278	C	T	T
905	rs3678308	4	120241830	G	A	A
31	rs3089514	4	123686862	A	G	A
419	rs3091114	4	129017764	C	T	T
740	rs3666032	4	131409009	A	G	GA
741	rs3696551	4	131618149	C	T	CT
906	rs3706432	4	132738971	A	G	GA
420	rs3023011	4	133639515	G	GA	-
218	rs3023026	4	140634063	G	A	GA
32	rs3023037	5	23743495	T	C	CT
742	rs3722968	5	25064528	C	T	CT
743	rs3670794	5	25068508	AT	GTA	-
421	rs3023040	5	32943077	T	CT	-
907	rs3714665	5	37385820	G	GA	GA
422	rs3023045	5	50863147	T	G	GT
744	rs3659745	5	51041036	AT	G	-
745	rs3713298	5	51194142	T	G	GT
746	rs3657668	5	66812086	T	C	CT
747	rs3672190	5	66812453	G	A	A
843	rs3657720	5	80572730	G	A	GA
844	rs3691937	5	101074114	C	T	CT
37	rs3023051	5	117479189	G	A	GA
33	rs3023057	5	125963636	C	T	CT
748	rs3685067	5	132834587	T	C	CT
749	rs3664890	5	135190462	G	G	G
424	rs3023060	5	142063375	C	T	CTG
475	rs3023062	5	145124169	A	A	-
425	rs3088741	5	145124441	A	A	-
426	rs3090833	6	18727440	C	T	CT
845	rs3023067	6	35949474	A	T	TA
750	rs3709317	6	49704741	G	A	GA
751	rs3716528	6	49740743	A	T	A
752	rs3659328	6	63635493	T	C	CA
846	rs3706583	6	63641367	T	T	-
753	rs3707041	6	63641402	A	A	AT
329	rs3090936	6	66965136	C	-	-
848	rs3152183	6	82717476	A	A	-
791	rs3704682	6	91992606	C	C	CA

792	rs3690102	6	100835585	T	T	CT
793	rs4137475	6	104904630	G	G	GA
847	rs3023083	6	106639546	A	A	-
430	rs3090025	6	109876793	A	A	A
754	rs3677586	6	110852971	G	G	GA
755	rs3707407	6	111404795	A	A	GA
849	rs3023092	6	127839602	G	A	A
40	rs3089737	6	145414611	-	T	TA
850	rs3090435	6	146590775	A	G	GA
45	rs3023116	7	33342591	CT	CT	-
606	rs3023117	7	34260777	G	A	GA
607	rs3023123	7	45323985	C	C	C
48	rs3023134	7	57286126	A	A	A
852	rs3711840	7	63737541	A	CA	-
608	rs3023129	7	65460296	A	A	A
908	rs3668498	7	69445393	A	G	GA
853	rs3704354	7	74666526	GC	GT	-
46	rs3090876	7	82744531	T	C	CT
854	rs3710192	7	85574011	T	A	AT
435	rs3090731	7	96549948	A	G	GA
480	rs3023154	7	113870672	T	C	CT
436	rs3089474	7	116219037	T	G	GT
756	rs3688884	7	124892578	C	T	CT
757	rs3682376	7	125149486	T	C	CT
438	rs3089174	7	128497838	C	T	CT
339	rs3023161	7	129084304	T	C	CT
758	rs3726791	7	139774847	A	T	A
759	rs3717254	7	140610603	C	C	C
820	rs3708255	7	143881399	T	C	CT
896	rs3023174	8	7812735	G	A	GA
118	rs3023175	8	11995469	G	A	GA
760	rs3700240	8	27809735	GA	GA	G
761	rs3665023	8	27811369	T	C	CT
439	rs3023183	8	35896398	A	G	GA
54	rs3088450	8	52979846	A	CT	-
762	rs3679837	8	60174269	A	G	GA
763	rs3686956	8	60210119	G	A	GA
610	rs3089230	8	71012852	C	G	G
897	rs3089636	8	71659444	C	T	T
609	rs3090460	8	74676357	A	G	G
51	rs3023194	8	91510521	T	C	CT
764	rs3711535	8	100556883	G	A	GA
765	rs3672284	8	100765637	GC	GC	GC
766	rs3724779	8	117005581	A	C	CA
767	rs3672332	8	117653778	T	C	CT
819	rs3696893	8	130456895	A	G	GA
768	rs3716232	9	16092806	C	T	CT
769	rs3694785	9	16382724	T	C	CT
876	rs3654109	9	32471104	A	A	GA
58	rs3023205	9	33547610	GC	GC	-
877	rs3672091	9	44035088	GA	A	GA
474	rs3023212	9	44086543	TA	T	GTA
878	rs3671494	9	59461009	GA	A	GA
611	rs3023215	9	65973933	A	G	-
441	rs3023216	9	71094035	C	T	CT
442	rs3023225	9	78134045	A	T	A
879	rs3695889	9	86124796	C	G	GC
770	rs4137954	9	98214467	A	G	GA
57	rs3090474	9	98298858	T	G	T
771	rs3688878	9	101293845	CT	C	CT
473	rs3089531	9	101742329	CT	C	CT
302	rs3023227	9	101742532	C	C	-
780	rs3707022	9	108156940	GA	A	A

781	rs3670181	9	111379601	T	T	AT
782	rs3682508	9	112519550	CA	A	CA
783	rs3657074	9	118619658	C	C	C
784	rs3700226	9	123845575	G	C	CG
445	rs3089912	10	20686036	C	T	CT
446	rs3090586	10	24286974	C	T	CT
772	rs3704164	10	40602386	C	T	CT
773	rs3705210	10	40612186	G	G	G
774	rs3696307	10	53431464	C	G	G
912	rs3702150	10	68557908	C	T	CT
777	rs3656551	10	81767829	A	G	GA
447	rs3090761	10	88548187	T	T	T
477	rs3090759	10	88548420	CT	C	CT
64	rs3089366	10	89110543	A	A	A
818	rs3722942	10	93074375	A	G	GA
314	rs3089906	10	97678673	G	G	-
778	rs3706590	10	107344215	A	G	GA
779	rs3660209	10	107394384	T	C	CT
72	rs3023249	11	11070846	C	CA	-
460	rs3023251	11	20853004	C	T	CT
909	rs3721297	11	21792473	A	C	CA
201	rs3023256	11	44299873	G	A	GA
461	rs3023258	11	54018188	C	A	CA
73	rs3088940	11	56359390	G	A	-
67	rs3023265	11	57576910	T	C	CT
799	rs3719410	11	67582900	C	T	CT
798	rs3719895	11	68481453	G	A	A
703	rs3709439	11	69031064	G	T	CT
797	rs3696966	11	69143793	G	A	GA
700	rs3659426	11	69692061	G	A	GA
701	rs3681957	11	69748839	C	T	CT
704	rs3707772	11	70267930	C	T	CT
68	rs3023278	11	71984385	T	C	G
910	rs3665064	11	82679842	C	T	CT
69	rs3089065	11	92981099	A	GC	-
463	rs3088501	11	94146845	C	T	CT
71	rs3023315	11	99324816	C	G	GT
911	rs3697014	11	112103534	G	A	GA
870	rs3725545	12	16183804	CA	A	CA
871	rs3686668	12	27493470	GA	G	GA
872	rs3724341	12	42468236	A	A	AT
483	rs3088822	12	53089104	A	G	GA
873	rs3690309	12	57606163	G	G	GA
75	rs3023347	12	57628517	C	A	CA
874	rs3700688	12	72297329	C	C	-
855	rs3703108	12	74842221	A	A	-
815	rs3662694	12	76915409	G	G	GA
875	rs3700106	12	77791482	G	G	GA
856	rs3697769	12	89115007	G	T	GT
465	rs3088800	12	99163402	T	G	GT
816	rs3726591	12	105250597	C	T	CT
305	rs3023378	12	113750831	T	-	-
466	rs3023377	12	113751107	G	T	G
81	rs3023379	13	18303792	G	A	GA
892	rs3717068	13	30844331	G	A	GA
79	rs3089102	13	44412430	C	A	CA
484	rs3089436	13	60912145	T	G	GT
893	rs3688361	13	80402371	G	C	GC
80	rs3023386	13	96657145	A	G	GA
813	rs3144879	13	101262056	C	CT	CT
485	rs3090063	13	104292074	A	T	AT

313	rs3023394	13	114216919	A	A	A
309	rs3023392	13	114837305	A	A	-
814	rs3708958	13	117780654	T	C	CT
785	rs3723026	14	7734219	A	T	A
786	rs3720966	14	10727081	G	A	GA
787	rs3712403	14	13327515	T	A	AT
788	rs3713871	14	24399404	C	A	CA
809	rs3724533	14	27108284	A	G	GA
789	rs3713838	14	27644295	A	C	CA
790	rs3685393	14	30183496	C	T	CT
811	rs3023409	14	48748605	A	G	GA
89	rs3023408	14	50023498	TA	CA	-
321	rs3021908	14	53097465	T	C	CT
86	rs3089070	14	67715311	T	G	GT
857	rs3705482	14	82835317	G	A	GA
87	rs3088599	14	84734491	T	C	-
858	rs3674616	14	98406482	T	C	CT
88	rs3090773	14	105296240	G	A	GA
812	rs3657504	14	118523326	C	T	CT
91	rs3023415	15	10765192	A	G	G
894	rs3088491	15	12374865	T	C	CT
492	rs3088634	15	19069275	T	C	CT
895	rs3662097	15	20153137	T	G	GT
96	rs3023416	15	31860737	C	T	CT
808	rs3669262	15	40601825	T	G	GT
92	rs3088488	15	42420068	G	A	A
859	rs3658370	15	51501719	G	A	GA
860	rs3701351	15	61455192	G	A	GA
93	rs3088506	15	71350770	C	A	-
486	rs3090719	15	74403438	C	A	CA
94	rs3088710	15	86619624	T	T	CT
913	rs3677860	15	87544783	A	G	GA
810	rs3717898	15	95702188	-	C	CT
101	rs3090912	16	6099904	T	G	GT
806	rs3667072	16	16273453	A	A	A
332	rs3089488	16	18511345	A	A	-
880	rs3695744	16	27296228	A	A	A
881	rs3718034	16	35665619	CT	C	CT
487	rs3089787	16	38850538	CT	T	CT
807	rs3663711	16	39824346	G	G	G
488	rs3090260	16	45609646	A	G	GA
448	rs3023243	16	58212994	T	C	CT
794	rs3654982	16	63555294	T	A	AT
795	rs3719654	16	70347393	CT	C	CT
489	rs3090908	16	72217393	A	G	GA
796	rs3695101	16	76223854	GA	G	GA
490	rs3090645	16	85879752	A	T	TA
455	rs3023436	16	88046993	T	A	TA
304	rs3023441	16	97552996	A	A	-
882	rs3674239	17	12357534	A	G	GA
104	rs3090500	17	25071190	A	CG	CA
884	rs3708501	17	30648071	G	CA	GA
43	rs3023110	17	32989898	T	C	CT
491	rs3088914	17	36695497	A	C	A
105	rs3023454	17	36882223	C	T	T
452	rs3023449	17	42120599	A	T	TA
862	rs3700023	17	52981854	G	A	GA
453	rs3023456	17	67310979	G	T	GT
202	rs3022791	17	71914187	A	A	A
863	rs3712928	17	78656292	G	G	-

805	rs3687592	17	81485883	A	A	GA
885	rs3710028	17	83573655	T	-	GT
312	rs3089323	17	83844781	A	A	-
864	rs3668190	17	91677947	C	C	-
605	rs3089544	18	5088109	G	A	GA
886	rs3696042	18	8912956	CT	T	CT
865	rs3655356	18	10241322	GA	G	-
887	rs3695261	18	21869777	C	T	CT
450	rs3023463	18	29943155	GT	GT	GT
803	rs3683689	18	33935868	A	G	GA
316	rs3089327	18	41908447	C	C	-
888	rs3657200	18	50109691	T	C	CT
109	rs3023468	18	63609445	G	C	GC
866	rs3657018	18	78455291	C	CT	-
889	rs3668347	18	80932716	T	GT	GT
804	rs3718427	18	88768623	G	G	GA
867	rs3665935	18	89757219	C	C	-
891	rs3669192	19	16913884	C	T	CT
800	rs3671678	19	17422426	G	A	GA
890	rs3658201	19	20407978	C	T	CT
111	rs3023481	19	20484928	GA	G	-
868	rs3710059	19	30660820	G	C	GC
336	rs3090951	19	44127717	C	C	-
801	rs3726430	19	44193090	G	T	GT
869	rs3713040	19	55813052	GA	-	-
449	rs3023498	19	60821445	G	GA	GA
802	rs3685993	19	60910395	C	T	T
602	rs3022803	multiple		C	A	-
7	rs3022821	multiple		G	GA	GA
401	rs3022823	multiple		G	G	-
725	rs4139354	multiple		GA	G	GA
418	rs3022994	multiple		T	A	CA
432	rs3023096	multiple		T	C	C
851	rs3723894	multiple		A	T	AT
776	rs3673999	multiple		G	G	G
817	rs3664582	multiple		G	A	GA
883	rs3668662	multiple		G	A	GA
914	unmapped	unmapped		A	G	GA
915	unmapped	unmapped		A	A	A
916	unmapped	unmapped		T	T	CT
917	unmapped	unmapped		A	A	GA
918	unmapped	unmapped		-	T	-
919	unmapped	unmapped		C	C	C
920	unmapped	unmapped		T	C	CT
921	unmapped	unmapped		C	-	CA
922	unmapped	unmapped		T	A	AT
923	unmapped	unmapped		T	C	C
604	rs3088804	x	99074052	A	GA	GA

Supplemental Table 2: List of PCR and SBE primers used for validation.

Assay ID	RefSNP ID	Chromosome	Nucleotide	PCR forward primer	PCR reverse primer	SBE primer sequence
713	rs3684370	1	15963812	AGGGAGGAAGGTGAGCAGTT	GCTGCGTATCACGCTCTCTT	TTTTTTTTTTCAGCAAGGAGTGTGTATGCA
714	rs3662163	1	32822284	AGCAAGCTTCCAGACATGGT	TTGCAACTGGTAGGAGCAGA	GCTGTATAAACTATGCCCCCAA
715	rs3704392	1	33320247	CCTCCCTCCCATTCCTAATC	TCATTCTGGAGTGCTTGGTG	GTTATTTATGCTCCAACAGTTATTGAAA
400	rs3090110	1	45907197	TGGTTCTTGACTGTTGACGC	GGTCTTCCAGAACCTGACCA	CTCAGCTCCTTCAGACCCTT
898	rs3667466	1	51794155	TGGCCTTGGGTGAGAATG	CCTGTCCATGTGTCCTCTGT	TTTTTTTTTTTTCAGGGACACACACATATTTTGCT
600	rs3022802	1	64979783	AGGCTGAGCTTCAAAGTTGG	TGTCAAGGGCATCAAGAAGTC	TTTTTCAAAGGGTATGAGAATATGGACTGG
602	rs3022803	multiple		CCAATCGTCACAAGTCAGTGAT	AAGAGATCATTGCCATACAGGAA	CACAAGTCAGTGATATTATAGCCAGC
899	rs3694327	1	79643870	GATTGATGATTGGTTTTATGTTGT	GGGATAAGTGTGAGTGGGAGTG	TTTTTTTAGGGCTTCACACATGTCAGTC
7	rs3022821	multiple		TCACTTGGTTCATCAGTTCAGG	GCCAGGGATTTATCAGAGCTT	AATAAAGTCCTTAGTGTGCATCAGA
717	rs3657283	1	93320420	TCTTCCTCTTCCCTGTCGTG	CCCAGGCTTTTTGTATTCTTG	CTTAAACCCCTGAGCCATCT
821	rs3704980	1	97461966	CAAAATGGGGTACAATCAAAGA	TGAAAGCTTCAGTTTCAAAGAAGA	GACTAGTCCACATTGGTGAGCA
822	rs3685188	1	110654171	CCAGTGGATTCACATCCTCA	TACCCACCAGCTTAGGGTTG	TTTTTTTTTTTCTTCTGTTCTGCTGGTTACAGCTC
721	rs3664018	1	121061603	TCACTGCCCAGAGGAACAAT	CTCAAGAGCGTGATTCACCA	TTTTTTTCACAGCCTCAGAAAGTCCC
720	rs3689749	1	121090273	CCAGACAGACAACAAGGGAAG	TCTCCCGTAGGACCATGTCT	TGGGGATGTTATGGGAGG
603	rs3022832	1	127102172	TGAGTCTGTGCTACCATGCCT	ATAAGAACTCATCACAGTCAGACCAC	AGACCACAGGAAACTCATTACTGTA
601	rs3090765	1	127691908	TTGTGACAGGGGATGTTCTG	TTTCCATAGCATTGGGGGTA	CCTAGATAAGTCAATACCACAAGTCCA
4	rs3022839	1	137257629	AACGTCACCAGACCACAAGA	AGGGTTTATTGACTTTTCATGC	GAGTAACATCACAGCCTTCG
823	rs3713224	1	154052109	ATTTCAGGCATGAGGGAATG	AGTGGGACAAGGTGATGAGG	TTTTACCAGTTTCATTTCCCATACTTCT
402	rs3022851	1	171617363	AGCATTGCTGAAACAGCTCC	TAATTGGTGTGTGTGTGGGG	GGGGTTATGTTGTATGTGTCTGTATG
403	rs3022871	1	174794383	GGATGGTTTTGGTTTGGAGA	AGAAGGGAATGAAAACGGCT	TTGAGACCAGATAAAGGACTTGAC
900	rs3716435	1	175899264	TGCCTTTCCAAGTTCTCCA	AGTTGCCTTTGTTGTGTCTGT	CCCACGTTCCAACACACA
824	rs3704926	1	184806157	CCTGTTTCCACCCCTATCCT	TCTGGGGCTATAGGCAAATG	TTTTTGCCGTTTTGTGACCGTCT
718	rs3713835	1	111883427	GCTTTTGTGTAATCATTGATATCTGA	TCTTGTTGCCTTAAGCAAATATCTC	GCAAAGAGGCCAAATCTGAA
719	rs3662850	1	111887865	AACAGGCCAATCAGCTTCTC	CTGGCAGTTGCCATAGTGAT	CCTCATGACTCATAGGAAAGATG
202	rs3022791	17	71914187	TCTGCCTTGTTGATGCTTTG	TCTCCATTCCTTGAGGGTCA	GCATGTAGGATGGCGAAC
401	rs3022823	multiple		CTGCACAAGAATCACAAGCA	TCCAAATGAAATCTTTGCCC	GGGTAGGGTAATTTTCTGACACATAA
712	rs3691476	1	15866460	TTGATCCCCAGAAACCACAT	CTAGGTTGGCAGGTCACTGAG	GCTTCTCTGGAGCTAGCATTC
716	rs3716254	1	93320412	TTCCTCTTCCTCTTCCCTGTC	CCTTGCCATTTTCTGAGCAT	CTCAGGGGTTTAAGAGTACTAGTTGC
722	rs3681847	2	11984102	CCTGCTTAACCGTGGCTAGT	TCCTCCTTGTCTGGGATACA	AGGATTGGATCAGCCATCA
723	rs4137557	2	11984314	AGGGACACTGGAGCAAACAC	CCTATCCCCTGAAGGTGTCA	TTTTTTTTTTTGGACCATTCTCCGTGTTCT
724	rs3664805	2	28174259	CAACATGCTTCATGGAATTTG	ATCTGCTCCTGAGCCCATTA	AAGATCATCAGGGGCCTG
725	rs4139354	multiple		GTTCCCTACTGAGGGAGTCG	AAGGCCATTGAGCTGTTGAG	TGATCTCATGTCTCCTCCCTTAG
10	rs3022883	2	37572959	CCCTGGACCCCTTTTCTCTA	AGCCTGGGTACTCACCACAC	TTTTCAGAGAGTTAATTTCCTCTGCAGA
726	rs3667376	2	45472806	TGAACTGCCCAGATTTTAATGA	CATGAAAGTTAGGACAACATGACC	GTTTGTGACCTGGTCTCTGTG
727	rs3681675	2	45990821	AAATGGTGACCAGCATATTTTT	CCACGTGCTGGAGTTACAAG	TCAGGTAATGGAAAATCACTCAG
404	rs3089489	2	51552294	GGGGGACACACACCATGT	TTCCTGTACAACTGGTTGCTTG	CTTTTGGATTAAAATGTCTGTGCT
405	rs3022887	2	68883673	TGCTCTGCATTTTGATTTGG	GACTGCATGTGGGGAGAGAT	TTTTGGGTGTTTCTATGATAACGCTC
901	rs4137272	2	79034808	GATGTGGCGGTAGCACTT	CCTCTTCCCTTTCATTCTTCTCT	TTTTTTTTTTTAACACTACTAAGAAGCCCTCTCAAGG
406	rs3022888	2	84403279	GGCTGTCCTTAACCCTCCTC	GCTTCCTGCCAGTTTCAGAC	TTTTTACAAAGCATCTTGCAAAAC
825	rs3669855	2	93586287	AACACCATCTTCCCAACCAG	TTGATTTATTGAAGCCCTAATTG	TTTTTTCAAGATGACTTTTGCCCTTGTA
902	rs3670817	2	94071001	GGGGGTCTTGGACTTTATCT	TTCCTCTCTCCATCCTCATTATTT	CTCTCTCAGAGTCTAAAGACAACCTGT
729	rs3692288	2	95029801	GACATGAAATTGAGAAGCACGA	GACCTCAAGGTCAGACACCAG	ATGCTTATCGCTGATAGATTGC
826	rs3663534	2	106743641	ATGGAATGAGTAGGCGGAGA	CCTCAGCATGCAGAAATGAA	TTTTTTTTTTTTCCAAACTCCAAGTTCTTCAGCC
827	rs3090608	2	117816539	GACCTGCCTCAAGAAAATGC	GAGTAAGGCAGAGTGGAGCG	TTTTTTTTTTTTCCGAGTGCTGACTCTGGGTT
18	rs3022895	2	119273525	GAGCCACCCATAGCTGAGAG	CACTCCGTCCAGTCAATGC	CACACGCCAACAATAAATGACT
828	rs3706262	2	119748062	CCAGTACGACCTGTGTGTGG	TATGTGGCCCGAGAATCCTA	TTTTTTTTTTTTCTCTGACTCCAGTCCCTCTGGG
829	rs3660910	2	133743791	AAATCTGACACATGGCTCGTC	GGCTGTATGCACGGTCACTA	TTTTTTTTTTTAATTTTGCTTTATGTGGCTGTCTATC
408	rs3022910	2	144973062	CTTCCCCCACACCTGACTAC	GGATTTTGGGAAGGAGAGGT	GGTTAATAAGGAAGATTTCAGGACC
830	rs3714631	2	151657817	GCAGAAAAGGGCACCTGTAT	GCCCACCCGTCTATCTCTCT	TTTTTTTTTTTTGAGGTGTGTTGGTGTGCACA
409	rs3089031	2	155762997	CCTGAGGACAACTCTGGAGC	AGCTGCAGAAGCATAGAGGG	TTTTTTTTTTTACGAGTAGAATTTACATGGCTCCTA
410	rs3022939	2	159674008	CCATTGGGGAAGGCTCTTAT	AGGGGGCAGAACTCTTTTG	TTTGTTAGGAGCTTAAAATGCCAGAC
730	rs3666331	2	167851555	GGGCTAGCTCCAGGTCAAC	TCAGCTTCCATAAGCATCCA	CCTCAGTGAACTGCACATCC
731	rs3674631	2	168354454	TTCCTGCGTTAGAGCCTACAA	GCTCCCTGCTGTCCATTTAG	TTTTTTTTTCAGGAACGCAGTGATTGTTC
732	rs3706063	3	12332500	TTGTTCTCCAAATAGCTCATGG	TCAGGCATGGGAAGATTTGT	TTTTTTTTTTTCATGTGGCTGGTTACCTCTC
733	rs3672323	3	12812319	ACGCCCCTTAAAGAAAGCTG	CTCCATTGCCCACCTGTAAT	TTTCTTAACTGTGAAGAACTAAACTGCAG
831	rs3676476	3	28110853	TCCCAGACTAGTCGCTAACTCA	CGTCTGATGTGGTGGAACAG	TTTTTTTTTTTTCACTCAGTAGGTCAGAGCAGGG
411	rs3022953	3	37599994	CCCCATTAAAGCTGCCTTCT	TTGACAAGGTATGCAGGGCT	AAGCTGCCTTCTTCATTTTCTC
832	rs3669022	3	42396732	TTCCCAGTTTCCCTCCTACA	TTGGTTGCTCTTTCTTACAGTCC	TTTTTTTTTTTCTTAGCCCTCCCCAACTTACC
833	rs3726717	3	52860548	AAAGGACTTGCTGCCATTTG	AAACCAACCCCACTGTGTTC	TTTTTCAGAAGCTCAGAAAGCATCAAG
20	rs3022960	3	65418042	TTCATGGGTGAACAAGAGCA	ACAATCAAAGCTGGCTCCAT	TTGAAACATGGAGACAAGGC
834	rs3689513	3	78626985	TAGAGAGCATGCACCCAACA	TGAAGTTCTTGCTTCGGCTA	TTTTTTTTTTTTGCTTTGGAGACATATTGTGTGGTTA

835	rs3669413	3	93426396	TCAATCAAACTTGTGGTCGTG	AGCAAACGTTGCTTGTTTCT	TTTTTTTCAAACTTGTGGTCGTGATAGATATTAG
836	rs3722681	3	109900215	AAAGCGCTAAGGTTCCTGTG	CAGGGCAGCACTTCTCTTTC	TTTTTTATGGACATCCGACTTCACTGTA
903	rs3672398	3	112220863	GTTTGTTTGTTTGCTTGGTAAGTATG	TGTTCCCACTGATTTCACCT	TGTGAATTAACACCAGAACAAGTAA
212	rs3022965	3	114011713	CAGGAAACCATTTCTTTCTGGA	TTTGAATGTGCCAACTCTTTTT	AAATAAAGTTCCATTTAAGGCTACA
904	rs3715748	3	119894667	CAGTACAGGAAGCAAGAGATTTTAG	GGAGGTGGGCTTTGAGAGTT	GCAGTTGAGGATGTGGACTCT
837	rs3691246	3	122466065	GTTCACCCTCAGACAGCTCA	TGTTGACAGCGGACCTAACA	TTTTTCAGTGCTTCTCAATAGCTCAGTTACT
734	rs3714750	3	126334813	TTTTCAGCTGGACGTTGTTG	GCTTCTAAAAGTGTCTCGTGGA	CATTAAGTGTTCCCCAACTAATCTC
412	rs3089257	3	130719076	CTCAGGGTCCTTAGCACTGG	CACACAGGCGACTTCGTTTA	GCCATTCGTCCCAGGG
413	rs3090381	3	140722145	GACATAGGTGGGTTTGGGTG	TAGAGGAAAGTCTGTGGCAGGTC	TTAAGTACTTGGGTATGAAGTTCTCAAA
736	rs3656469	3	143498147	TGGCACATTAAGTTTGCACAG	TGTTGCTTGCATTTTCTTTTTC	TCTCCTGCTGGGCTGTG
737	rs3706436	3	143786333	TGGAGGAAGTGTGTCACTGG	TGCTGGACAGGTAGCTCAGA	GGCTCAAGCTAGGCCCAG
471	rs3022975	4	8073146	GTGCACACACCACAGCTTCT	ATGACTTGGCTCTGGGAATG	GCTCTGGGAATGTGCTTTTC
414	rs3091112	4	8073150	GTGCACACACCACAGCTTCT	ATGACTTGGCTCTGGGAATG	CAGACATACTAATATGACTGTATTATATATCAATATGA
738	rs3711350	4	26559488	TTCTGAACAAACAAGATGATTGC	CCCATGGTCTACCTTGGACA	TGACAATTTCTCACATGGTATTAGATC
739	rs3684156	4	27375172	GGTCTCCATATTTGAAGGGACA	CAGAAAAGGTTCTCCCAGAAA	GGCTGTCTAAGCATGATCTGAA
838	rs3665192	4	39271744	GCTGGAATTATTGGCAAAGC	TTTCTTTCCTGCATGCTGTT	GCATATCCCAGCAGCATG
839	rs3667625	4	54215713	TTTCATCACTTCTCCCAAGACA	AAAGCATCACAATGCAGCAG	CAGACATTCTTACTCCATCATCTCC
840	rs3694308	4	62440427	GCACTCAGAACCCAGTCACA	ACTTCTGGCCATCCACAGAG	TTTTTTCAAGTATATGGACTTGGAAGACAATG
415	rs3022979	4	75874108	GGTGCTATCTTGAAGCCCAG	GAAGAGGGCATCAGTGTTCC	TGAGGGCATTGGTTTCTTTC
841	rs3707288	4	85885441	AGCAGCCTGTCTGCATGG	TCCTGCTGTTTTCAGCTCCT	TCACCTGCTCGTATTCCTGGA
709	rs3713394	4	86423755	AACAGGGAATCTGGGAAATG	ATCACCACGAGTGCATGTGT	GTCCTTGCTAAGCATGCATG
707	rs4135993	4	88399014	TGGTAGGGAATCCAACCTTT	GAGAGGGGCATTGAAATTGA	AATTCTAGCCCTGCCAAACAT
708	rs3714181	4	88415879	TGAGTCCCCAGAGCTTTTCA	ATGGCTGGGTTTTGTAAGCA	ACAAGCATTATAGTTCATGGCTTTC
705	rs3680265	4	88774811	TGCATACCACGTGTGTTCTTT	CCTGCTGCAAAAGACTCCTC	CATTTGCCTGCGAATTTCA
706	rs3696308	4	88775106	GGTAGAGGTGGAAGCTTTGC	AGATGGCCTCGAACTCAGAA	GGGCAGTGGTGGCGTA
416	rs3090804	4	92499971	CCATTTTGGCATGGAGTTTT	AAGGCTTAGGCCTATATTCCTTTA	ATTTTTGAAGGCATCTAAAGTTTTT
417	rs3088455	4	99679889	TAGAAAATCTGGGCTGGGAA	CTTTCAAGTCAGGGCAAAGC	TTTCACTTTTCTATCCCAGATGCC
478	rs3022989	4	99679747	TAGAAAATCTGGGCTGGGAA	CTTTCAAGTCAGGGCAAAGC	ACAAAACTGCCAGTTGCTTC
418	rs3022994	multiple		AAGCAATGCTGAGCTGTTAGC	CAAGCATGTGAGTGAGTGCC	TTTTTTTTCACAGCAGAGACCAATCCAG
842	rs3658845	4	115459278	AACCCTGACTGTCCTGGAAC	TCAAACTTCCTGGCTCCACT	TTTTTGCAGGCAGATCTCGGAGTT
905	rs3678308	4	120241830	TCACTTCCCAGAGGCACA	TGGTTGAAATACCTCCCTCA	TGGAAGGTGTGGGTCCTG
31	rs3089514	4	123686862	ATGTGGTCTTGTTGGGGATG	TCCAGCCTTTATCTGACCAGT	TTTTTTTTTAGAGGCACTGCCCCA
419	rs3091114	4	129017764	TCTTGTGCTGTGCCTAGTGG	TACAGACCTCCCTCCCTCCT	TTTTTTTTTTTTTTAGCTTGTGAAATCAACCATGTC
740	rs3666032	4	131409009	CGCCGGGACTTTGAAGTTA	TGCAGAAATCAGTGGCTTTG	GTGAAGGACAGACAGACAGACAG
741	rs3696551	4	131618149	CTGAGCCATCTTTCCTGTCC	TGTGAGCAGAACCTTCCCATA	CTGTCCCCGTAGACTAGACCTT
906	rs3706432	4	132738971	CCAGGAGACCCAGGAAAG	CCTCCACCACCAGTTCACAAA	TTTTTTTTCGTTTCCTAGCTCAGATACTTGGT
420	rs3023011	4	133639515	TGTGCACCATATGCTCAAGATAC	AAACACTTGCTTCACAAGAGAGC	TTGGTTGTCAGCCCTCATAC
218	rs3023026	4	140634063	GCAAACATTTCCATTCTCAGC	CACATTCACGCTTATTCATACACC	TTTTTCCACACCTCCACTATTATAAAGC
32	rs3023037	5	23743495	CTGGGTCCCATGATGCTTTC	GTGTGGATCAGGGGACAAAG	TTTATGGTGCCAGAAAATCAAC
742	rs3722968	5	25064528	CTCCAAAGATTTCCAGACAGC	TCAAAGGGCTAAGGTCATGG	TCTCTCCTAGCTCATCCCATG
421	rs3023040	5	32943077	CTTCCAGTGCGTGCAGAATA	AACATCCACAGGCTACTGCC	GCTAAAGGCAACATTTACTAATGACA
907	rs3714665	5	37385820	CACTCTGCTGTCTTCTGGACT	GCTGTCCTGCCCTTTACCA	TTTAAGAATCTAAAGCATCCGGCA
422	rs3023045	5	50863147	GGCAGTTGGTGAGTCATGTG	AAACTGTTTCCTCCTTCAGGG	TTAACTTGGAATATCAGGCTTTCTTT
744	rs3659745	5	51041036	GGGAGGGGGACAGAAAAAG	CATGGTCAGGTTCTTTCTGGA	TGTGTACCCCAATAGTTATACTCAATAG
745	rs3713298	5	51194142	ACCCACCTGAAATGGACAGT	CAGATAAATTCCTGTTGGTGGT	GGTAGCTGCTAAATTAATCTTCAAGAG
746	rs3657668	5	66812086	TTTCCGAGGGAAATGGAAAT	GGGAGCATGTGCAGTATTGA	GCGGCTTCATTCTCCATCT
747	rs3672190	5	66812453	GAACAGCTGTGTGGCTTTCC	ACACTGGAGAGAGGCCTGAG	CTCTCCCACCCTGCACC
843	rs3657720	5	80572730	TGAAACCACCTCAAAGTCCA	TGGTAGAGAAGCCCAACCTT	TTTTTTTTCCTTGTAATAAGCCACAGCATCT
844	rs3691937	5	101074114	GCCCTGGAACACTGCTAAGT	AGCCACCACCCAAGTTTCTA	TTTTTTTTTGCCTCTGTTGTGGCTAATCA
37	rs3023051	5	117479189	ACCAACAGACAAAACTGCCC	AGTGATGACCCTGTGGTGTG	TTTTGAGAATGAAATGAACACCAG
33	rs3023057	5	125963636	CCAAACTGTTGCAAGGACAA	GACATAGACACAGACACCCCC	TTTTATCGAGGTGCCTTTTGTC
748	rs3685067	5	132834587	GGAACCTCAGTTCCACCAAA	GTGCAGGTGAGTGGGAGTTC	TTTTTTTTTTTTAGATGCCTCATCTGATAACAGG
749	rs3664890	5	135190462	GCAAAGCTGCTGGACAAAAT	TGCACATAGAAGCCAGAGGA	AGAAATAAGGGAGAGGGAGGG
424	rs3023060	5	142063375	TGAGGTCTGTGTTTCCACCA	CCTTGCCCTTTACACTCAGC	TTTTACGTGGCACAGACCTCAG
475	rs3023062	5	145124169	GCAAGCCCCAAAGAATTGTA	AAGAACAGCACGTTGGGTTT	CCAAAGAATTGTAACTTGCCC
425	rs3088741	5	145124441	GCAAGCCCCAAAGAATTGTA	AAGAACAGCACGTTGGGTTT	AACTACGTGAGGAAGAGAATCACA
914	unmapped	unmapped		TGGTCCCTGAAAAAGGTGAA	GCATGTTTTGCTGTGGGTTT	CTATACTTAACCCAGTGTCATTTTGAAA
426	rs3090833	6	18727440	GGGGAACAATTTGACTGTCTG	AACCTGTGCTGATCCCTTTG	TTTTTTTTTTTGAAAAAATTGCTTGTACTGTAGTCT
845	rs3023067	6	35949474	GCGATATGCTATCATTTTTGCC	CACCCATAGTGTGTGCTTGC	TTTTTCTTGCTGTCCTTTTGAGCTGAG
750	rs3709317	6	49704741	GACGTCAGCCCCATTCTAAG	TCGTCATTCTTTGTCCCAACT	CTGCAATGAACATCACAGAGC
751	rs3716528	6	49740743	CAGAATGTGAGAACAAACTTTCTCTT	AGAACACATGCACTTGCACA	GCATAAAAGTGCTCACATGTGTG

915	unmapped	unmapped		CAATCCCTGACCCTGAAGAG	CAGGCCCATGTATCCTCTTTT	AAGGCCCAGGACCACA
846	rs3706583	6	63641367	GCTCCAGGAGGAAGCATTAAG	TGTTCCTGGTACCACAAGGT	CTTCAGTCTACAGTGGTATGCTATCAG
916	unmapped	unmapped		AGCTGAGGGAAGTAATGCACA	TGCTTCTGTGTGTCAGATTGC	TCCGAGCTCAGTGTGCTGCTCT
848	rs3152183	6	82717476	CCTTCATTGTCACCGTTCCT	TGTCTTGGAAGAAGCCCAGT	TTTTTGGTAGGGCTTGAGCTATGACTAGTG
917	unmapped	unmapped		GGTGTCCTTGGGCTAACCTT	TTGTGTAGCGATTGGAGCTG	TTTTGGTGGGCGGACAGAGC
791	rs3704682	6	91992606	CCCTGGTTCTCATGCTATGG	CCTGAGTACATTCCCTGACCA	ACCTGAAGCTCGCTGATTTG
792	rs3690102	6	100835585	GCCTAGAGCTACTGCGAGGA	GCTCAAAGGCAGCCTGAA	GTGGAAATGTACCTGCAGAGC
918	unmapped	unmapped		CAGAGCTGGTGTCCTTAGTTTT	CTCTCCTTCCGGCACAAAG	AAACCACCTCCAAACCTAGAAAGT
793	rs4137475	6	104904630	CCCCAATCATACAACTGTGC	TCATCATGTATTTGGTAAGTGTTGG	TTTTTTTTTTTTTTAAGTCACAGATGATTCTACCCAATCT
919	unmapped	unmapped		AGGAATAACAGGAGGGCTTG	GCCATATCCATCTAACTTGTAAAATAA	CAGGTGCCACAGATAGTGAACA
847	rs3023083	6	106639546	CTGAGAAGCCATTAGCCCAG	GGTGGCAACACAAGGAAAAG	TTTTTTTGGAGTAGCAGGTTGCTCAAGAG
430	rs3090025	6	109876793	TGGATTATAGTAAGGTCTCTTTGGA	GCCAAAAAGAATGCCTGATT	TTTTTAGTTTATAAAATAACATTATAAAAAGATGGG
754	rs3677586	6	110852971	AACTTGCCGAATGAACTCCT	TCATGAAGGCAATAATCACACA	CTAATCTGTAGAGAGGAAGACCACA
755	rs3707407	6	111404795	TTGTAACACCAATCTACAAACCAA	TGTGCTGTCCTTTTAGACATTCA	GTCAGCAGCTCACAAAGTCCT
920	unmapped	unmapped		TCAGATGAGAAAAGAAGGAAGTACA	GGGGAAGATGAGTGTCTGTGA	CCCTGTGGAAAGACCTGATAAC
921	unmapped	unmapped		TTTCTGTGGAGACTTTGAAGCA	TTGCAGCCAACACGTTTAAG	TTTTTTTTTTTTCTGGCAGAAAGGCCCAATGTTTAA
849	rs3023092	6	127839602	TGTCTGTCTATCATCTGTCTGTCTG	TCAAGTTGACAGCATTAACTACCAC	TGACAGCATTAACTACCACAGCC
922	unmapped	unmapped		TTTTGTTTCTTTGGGGCAGT	GATCCAGTGAGAGCCAGACA	TTTTTTTTTTTAGAGCCAGACATAGTAGCACACGGA
923	unmapped	unmapped		TTAAATGTGGGTGCTGAGAAT	GCCAGTAGGTGAGTAGAGGTTTG	AAACCCCCACCTCCTAATTCT
40	rs3089737	6	145414611	CCAGAGATTCCGCTGTCTGT	TTGAGGTGCTTCGTGTTAAAAA	CCTAGTAACTTCAGTTTTTAAAAACTCC
850	rs3090435	6	146590775	CTTTTCAGTACTGGCCTTAGTTG	CAATAGTCTTCTGAGTGATGGGA	TTTTTTTTTTGGTACTATCCCTGGCTTTTCACA
820	rs3708255	7	143881399	TTGTTCACCCATGTCTGTCTG	GCCAAGAGCTGTCCAAGGTA	TTTTTTTTTTTTCCTGTCCCTTCACCAGGG
45	rs3023116	7	33342591	GGTCTTCCTGTGTAATTTTGGG	TTTGGGGCTAGTTGTCATCA	TTGATATCAGTATTTTGGAGTAAGAGC
606	rs3023117	7	34260777	TTTTTGAGGCGGAGTCTAGC	TGGCTAAAGTTGCCAATCAA	CAATAGATATTTAGATGTTGCTATTGTTTATCTAC
607	rs3023123	7	45323985	TGGTGCGGTGATACACATCT	GTGAGCATTCCACCACCTTT	AGCACTCGGTAGGTAGAGGTAGA
851	rs3723894	multiple		TCCTACCTCCATCTGCAGTAAAA	GCCAACAAGATCCAGGAATG	TTTTTTGGCTAGTTCATAGAGTATCAGAAATGTGT
48	rs3023134	7	57286126	GAAGTCAGAGTGGCCTCACC	TGACCCACAAAGCCTCTTTT	TTTTTTTCAGAGTGGCCTCACCAG
852	rs3711840	7	63737541	GGCTTGAACACCTGTCCCTA	CAGGTACAGTGGAGGGGTTT	ACCACAGGGGCCGTCTAC
608	rs3023129	7	65460296	CAGATCGTTTCCCCCTTTATATC	CTGAGGGTCTTTCAAGAGAGGTT	ATGGAGCCATTCACACACAG
908	rs3668498	7	69445393	GCACACAGAATAGCAACTGGA	GTGGTCCTCTGCCCTTTAC	GCCAACCACAATGCCA
853	rs3704354	7	74666526	CAGCCCACGTGTGTTAAACC	TCAAAGGAAAAGGTTGGAAGA	CCCACGTGTGTTAAACCCAC
854	rs3710192	7	85574011	CAGGTCATCTTCCAAGTCATCA	TGCTCCATGTTAAGCATTACAATTA	TTCTCCATGTTAAGCATTACAATTATGACTA
435	rs3090731	7	96549948	CACTTGTTGCCACTGAGCAT	CAGAAATCGGGCGAAAAGTA	CAGGTCTCCAGCTGAAAGC
480	rs3023154	7	113870672	TCCCCCTGGATCTTCTCTTT	GGGATGTGGACTCTGAAGAAA	CATGGCTCTCGCCTTCT
436	rs3089474	7	116219037	CGGGTGTCCTGGATGTTTAG	CACTAGGCTTTCTTGCCAGC	TTTTTTGTTTAGAATGTTGCTCTTAAGGGTTT
756	rs3688884	7	124892578	CCAGTGCTGCTCTGTGACAT	AGGAGAAGGTGGGGAGAGTC	TTTTTTTGATGGGTGGGTGTGTCAGT
757	rs3682376	7	125149486	CTGTTGCTGAACCATGCACT	TGGTGGGAGATCCTTAGGTG	CCTTTCTCCTATCTTTACTTCCTAGC
339	rs3023161	7	129084304	CTAACAGGGAACAGGCAAGG	GCTGTAGCCTGGTTCAAGGT	CACCACTTGATATAGGAATGTACACTT
758	rs3726791	7	139774847	AGAGGAGAGTCCAAGCCTCA	GTCAGTGGGTAGCACGTCTG	TGATCCTGTCTCAGTTCCACC
759	rs3717254	7	140610603	TTTTGTTGGCATCTCCTTTCTT	ACCCGACCTCCAATAACAGA	AAACCCAAGGCAAGACTCC
896	rs3023174	8	7812735	TTGCCCCCAAAATTAACAAG	CAGCTGGTCTGCTTTTTATGG	TTTTTTTTTTTGAAACACAGTTAACCAAACCTCAGA
118	rs3023175	8	11995469	CAATGCACAAGTCAGCATCA	GATGAGGCCTGACAAAGCTC	TTTTTTTTTTTGCACAAGTCAGCATCAACGCAT
760	rs3700240	8	27809735	TTTTGGTGTTGTTGTTGTTGTTC	CCCTCCCACACCAATTATCA	CCACAGCCATCTGGGAA
761	rs3665023	8	27811369	GGCTCACACATGCTCAGAGA	GTGGTCCAGGCCAGATCTTA	TGGCTGAGGGACTTGTGC
439	rs3023183	8	35896398	AGCTGAACCCTGAAAACAGC	GCCCAGACATTCCCATGTTA	TCTCTGTGCCCACAGCCA
54	rs3088450	8	52979846	CACGGTTGTGACATTTGAGG	GAAGGCACATAAACGGATGG	GAGGAAAACTCTTCTTAATAGTGGAAG
762	rs3679837	8	60174269	GACCCTGAAATCTGCTCTGG	GGGGACTAACTGTTGATTTGAA	ACTGCCAGTTCATGACCTCC
763	rs3686956	8	60210119	AGGGAGCAGGCTGACATCTA	ACCCCAGGGGGTGATTATG	ATGTGACAGGGAAGTTGGC
610	rs3089230	8	71012852	GCTGCCCTAGCTCTTTGAAGT	TCTGATGATTCTGGTGGCAG	AGCCAGGGCTACTGCATAAC
897	rs3089636	8	71659444	GGGAACAGGCTCACAAAGTC	CCCAAGGCACAGCTTAGTTC	TTTTTTTGAAGAGCCTTTCTAAAGAATGTTTTAAGTATA
609	rs3090460	8	74676357	AGTGTTGTGCACCCTCACAG	CCAGAACAAGCCAGTCCACT	TCCATTCCAGAACTTGTCCATA
51	rs3023194	8	91510521	TGGAGGGAGACATTGAGAGG	GGTCCTTGGGAGGTGCTTAT	TTTTTTTTTTTTTGCCTTTAGGTTTCATGC
764	rs3711535	8	100556883	CCCTTCCATCATGTTCTTCG	AACAGTCATGCATCCCCTTC	TTTTTTTCCATTTGTTCCAGAGGCA
766	rs3724779	8	117005581	TGTAGCAGGAATGGCAAGTT	CCTCCACCTGTGTCCTCATT	AGAGGACCTAATTGAGGACTGC
767	rs3672332	8	117653778	TAGCAAACCTGGGACTCAGC	GACCAACCCACAAGAGCCTA	ACAGTGAGGAGGGGCCA
819	rs3696893	8	130456895	AAGGCCAGGCAAGTCACA	GTGCTCGCACATAAAGCAAT	TTACAAACAGACTTGCCCTGAAAG
768	rs3716232	9	16092806	CAACACCCCAGGCTTTACAT	CCTGCTGCACTTGTGTCTGT	TTTTTTTACACATCACAAGGCCACCTA
769	rs3694785	9	16382724	TTCCTCAGAACAGTGGAGAGA	CCTGCACAGCTCTCAAGATG	TTTTTTTATACGACTTAGCTACAGTCCCTGG
876	rs3654109	9	32471104	CTTTGAAACTGCCCTTTTGG	GAGCCTTCCCTGCCTTACAT	TTAGTTGCAGGTGAAAGTTCACAA
58	rs3023205	9	33547610	GATGACGCTGGTGGAGACTT	AGTGAGCCGTCTAGCACCAT	CCAAGCCTAGCCCCAGTA
877	rs3672091	9	44035088	GCCCTCATCCCTCACCTACT	TCTGCTCTGCTGGAGAAACA	TTTTTCCGGTCCGATGGTGGGCATGG

474	rs3023212	9	44086543	TAATACGACTCACTATAGGGAGATTGCTTCAG CTTGGTGCTTA	AATTAACCCTCACTAAAGGGAGAGCGACAG GGTCAACTGGTAT	TGTGATGCTTTTCTCCTCCTC
878	rs3671494	9	59461009	GGTGGCTGGCTTTTTCAGTAG	CGCTAAGGGACAGTACAGATGG	TTTTTTTTCCGCTCACTTGTCCCATCTC
611	rs3023215	9	65973933	CAGCAGGAAGTCTTATAGGAGCA	AAGCCATCTTAAAGCAGGATACC	AGGAAGTCTTATAGGAGCATGCAC
441	rs3023216	9	71094035	GGTCCTCCGTCATACCCC	AGTCAGCTTGCTTCACACCTC	GGCAAGAGTGGTGAATGGAC
879	rs3695889	9	86124796	TTCTGACCTGCTTGCTTTCC	GGGCCTCATTGCTCACAG	TTTTTCAAGTTAGAAGCATTGCCCTC
770	rs4137954	9	98214467	CACAGGAAGCCCATTAGTGC	GGCACGAAGTTCCACTCCTA	GCTCTGGTAAGTTTAACACACTCC
57	rs3090474	9	98298858	CCCTTCTTATCTTCAGCCCC	ACAGAACCTGGAAAATGAACTG	TTTTTTTCCACCAGGCCAGCTT
771	rs3688878	9	101293845	TGGGATCATGTCCCTAGAATG	GCTTGGTCTGCTCACAATCA	GAAGGAAGTTAGGGTTATATGTAAAGC
302	rs3023227	9	101742532	TTGGTGTGCATGGGAACTTA	TGCTGTTGGACACACCATTT	TCAAGAGGCTGCTTCCAGAA
473	rs3089531	9	101742329	TTGGTGTGCATGGGAACTTA	TGCTGTTGGACACACCATTT	GGAAACTAAATGTAAACTTGGGAGTA
780	rs3707022	9	108156940	GGACAAGGACTTGGCTCTCA	CTGGTTTGTACCAGGGAAGG	AGCCCCTATCTGCATTGAGA
781	rs3670181	9	111379601	GGAGATGTGAGGCAGGTCTT	CCCAGCCCAGAGCTATTACA	TAAGTCAACTTTATAAGTATTTATCTGCATTTACT
782	rs3682508	9	112519550	CTTGGCATTGCTGAAGAGTG	CGTGCTCTTTGCCTTATTGC	GCGAGCCTTTCATCACAG
445	rs3089912	10	20686036	AAAAACAGAAAAAGGCAGTGACA	AACAGGGCACCATAGAAACAA	TTTTTTTTTTTTTTTGAAAAAGGCAGTGACAAAGTATG
446	rs3090586	10	24286974	ATCCCACCAAAGTGAGAACG	CTTTCTCCTGGCATCTCCAC	TAGCGCACAGTGCCAGAA
772	rs3704164	10	40602386	TTGTTGTCATGTGACTGATGGA	GGCAGAGGTGGGTGTAAGAA	GCTGCCATATGAAGATCTCCTCTA
773	rs3705210	10	40612186	CAACCCAAAGAACACACAACC	GAACCGCAATGGCTAAGAGA	CCTATTAGAACGTAATTCTGCTTTCTAATA
774	rs3696307	10	53431464	CCATGCTACTGCACAAGACAA	CCCCTTCTTTCTCCCATGTT	AGTGCAACATGGGAAATATGA
448	rs3023243	16	58212994	GAGGCTAGTCGGCAACTGAC	TCATGTACCAACTCCCACCA	TTTCGGCAACTGACTTTGGACA
912	rs3702150	10	68557908	GCCCCCATTTATTCTCAACC	GGCAGGAACAGAAAAGGAGTG	CAGGCGGGCTAAATTCA
777	rs3656551	10	81767829	AGAAGATGTGGGCCATTGAC	CATGGGGGAGACTATCCTCA	TTTTTTTTCTCCAATGCCTCACTGGATGCTAC
447	rs3090761	10	88548187	TGGTCCTGGCTTCTGTCTTT	GCTGACCTCATTTCTCAGCC	TTGCAGTCTTTATTGATGTAAAGGAACA
477	rs3090759	10	88548420	TGGTCCTGGCTTCTGTCTTT	GCTGACCTCATTTCTCAGCC	CCACTTCCTCTCTCTGATTCTGA
64	rs3089366	10	89110543	GGAAACAGCAGTAAGGAGAACG	TGAGTTCTTCCTCCCCTGAA	GTAAATGGCTTCTTCTGCACTTT
818	rs3722942	10	93074375	CCACCGAACTCTTCTGCTTC	CTTTTCCCTGTGCACATCCT	GCTGCCCATTCTCACCT
314	rs3089906	10	97678673	GAGACAAGACAAGGCAACCC	CTAGACATCTTCGGGGTGGA	CAACCCAGCATATCCCAC
778	rs3706590	10	107344215	CGATAGAAAAGTCCAAAGAATGAG	AGGGTTCTGTCCTCTTTGTCC	CTATCCCGGAAGTTTCAAATAAC
779	rs3660209	10	107394384	CAAGAGTTTTCATGGTTGTAGCA	CAGTTGACAGCAAACACTCACT	GTCAATAGGTGAGAAAAATATCAGACTG
72	rs3023249	11	11070846	CTCTGTGCCCTTTTGACCTT	AGAGACAGACGCCTCTGGAA	ATGGAGGGACACCTGGG
460	rs3023251	11	20853004	CCAAGGACCCTGCAATTCTA	ATCGCAGCTGGCTAAAATGT	GAATGAGTCAGAGGTCCTTCATTAG
909	rs3721297	11	21792473	TCCTGGCTGAGGTTGATG	GAGCTTGCCTACAGGGAGTG	AGGCAGGTCCATGCAGG
201	rs3023256	11	44299873	TTGATGACTTCTTGGAGCTGAA	CCTTTGAGGGGACTCCAGTA	TTTTTTTTTTTTTTCTGGGAGAGGTAACTGCTAACT
461	rs3023258	11	54018188	CTTAGCCTAATTTGGCAGTTCCT	ATTCATGCACACATACAGGCAT	CCTCCCATTATTTTTTTTTCTGAG
73	rs3088940	11	56359390	TCCAGAGTCACACCAAGATCA	TATCTGTGCTGGGACCATCC	TTTTTTTGGGTGTTCTTTCTTTGAAAGTGTTA
67	rs3023265	11	57576910	CCTCCTGTCTAAATGAGAAGGTGTA	AGGTGTCCCCTGTCATTGTC	TTCCTTATTTTGCCAGTCTCCTTACT
799	rs3719410	11	67582900	CAAGAGGATACAGTAGCCAGCA	GCACCAATTAGCCTGTTGTTT	TTTTTTTTTTCAATGAGCTTTGAATTTCTGCTAATAA
798	rs3719895	11	68481453	TTGGAGAGGGAAAGATGTGC	ATTCAATCCAGTGCCAGGAG	ACAGTTGTGTTCAAATTGGCTC
703	rs3709439	11	69031064	CTCCTGTCTGGCTACCTGGA	TGAAGTTGTCAAGACCAGCA	TCAGGTAAATACAAAGGACTTGCTT
797	rs3696966	11	69143793	AGACACAGGCAGGACCTCAC	CATAGCAGAGGCCCTAGGAA	TAGGAGGGAACGGAGGC
700	rs3659426	11	69692061	GGGGGCACCTCTAACAAGA	CTGGCATACAGTCCCTCCAT	TTTTTTTTGGAGTCAGACAGGAATGGAA
701	rs3681957	11	69748839	GATGTGAGACTGGGGGAGAA	GCTGCTCAAGTTGGCCTTTA	TTTTTTTCTCAGGTACTGAGTGAGTTCCTAGACT
704	rs3707772	11	70267930	CACTGGAACGATGGCTACCT	GCTGGTCTTCACCCCTTG	CCTGGAGGATATCGAGCG
68	rs3023278	11	71984385	TAGGGAACTGTCCCTGCCTA	ACTCCTGTTCCCTTCTGGCT	TTTTTTTTTTTTGTGCTTGCTGCTGTAATTGC
910	rs3665064	11	82679842	TATGCCAACACTCCACTCATTC	AAGGACTAAAAACCAGGGAACA	TTTTTGCATTGCTAATTGTCCTGTGG
69	rs3089065	11	92981099	CAGTGAGATGGGAGAGACGTG	CCTGATTCCTTGAAGTAAGGC	TTTGTTAAAAGGAGGAAGAGAAACATA
463	rs3088501	11	94146845	TCACAGTAACATGGGAGCCA	TGTGATACAAAGCCAGCCAG	ATGCTAGTAGGAAGACTCTGGAACTA
71	rs3023315	11	99324816	TGGAAGGAAGTAGTTAGGATGACTT	CCCATAACATTATTAGGAAGGTGTC	ACTTGTAAGGGTTAGGGTGTTT
911	rs3697014	11	112103534	TGAGTCCAACAATTCTCAAGGT	TGTCTTACTTTACTTCTTCCCTGTTC	TTTTTTTTTTTGCTTAAGCTTGTGTATCAACTGCCT
870	rs3725545	12	16183804	TCGGTGCAAGACATTAAGTGA	GGTTTGGATCACCAATCACA	CCTGGTCTATGTGTTCCCAGG
871	rs3686668	12	27493470	TGTCGCTTGTCTATGCCTGA	GATCTTTCATTTTCTTGGGGATT	TTTTTTTTTTGAGTCAGCCATGTATGTGATAGTGT
872	rs3724341	12	42468236	GCAGCATATACAGTGCCTTCC	CAGTTTTCAGGTTTGTGGACAG	GTCATGTTATTAACTATGGAGAAATAGAGAAAC
483	rs3088822	12	53089104	ATGCTGTCAGGCTTGTGTTG	CGCATTCCTGAGCTGAATTT	GTACCTGTGAGTATTCAGTCAGCA
873	rs3690309	12	57606163	AAGGTCCTACCTGCTCCTGTC	ACCTGCATCGGTCTCCAG	TGCTCCTGTCCAATCTCCA
75	rs3023347	12	57628517	TAATACGACTCACTATAGGGAGATGTGCCTTT GGTTGTTTTGA	AATTAACCCTCACTAAAGGGAGAAGGTGTTG GTGTTGGTGGAG	ACATCCATATTTACAAGGTCATAGAA
874	rs3700688	12	72297329	CCAGCCCCCAAAGAGTTATT	CCAGTTCAGTTCCCATCACC	TTTTTTTTTTTTGATCTCTTGGAACTGGAATTACAGAC
329	rs3090936	6	66965136	AGAAACTGATGTCAAGGGGC	TGAAAGTAACCTGGCCATCC	TCAAAAGGGACAAAAACTACAAAG
855	rs3703108	12	74842221	AAGAGACCCTGTCTCAGGAGAGT	ATTCAATCAACCGTGCGTGT	TTTTTTTTTCCTGTGACCTTCTCATGTATGC
815	rs3662694	12	76915409	CCCTCTCTGCTGGTACTGAGA	GCATGGCAGGGAGATTAGAA	TGTACACTTGCTGAGATACTGTGC
875	rs3700106	12	77791482	CGACATTCTTTCCGCTTTTC	AGCACTGGGGTCAGCAATAC	AGGCGGGACTTCTTGCTTA
856	rs3697769	12	89115007	CTTCAGTTGCTCCAGGCTTC	ACTCCCAACTTCAGCCACAC	CCTCACATGTGGACTCAGGC

465	rs3088800	12	99163402	CGACACCCCCACTCTTAATG	ATTCAAGGAATGCTGCCACT	TAACAACTTTAGTTAACTAGAAATACTAAGTCTTGA
816	rs3726591	12	105250597	ACCCAGATGCCTGAGAGATG	TGGCTTAATGCTTTGGGGTA	TTTTTTTTTTGGGAGGTCGGTATTAGGAGAC
305	rs3023378	12	113750831	TAATACGACTCACTATAGGGAGAATGCATGCT GTTGTTTTGGA	AATTAACCCTCACTAAAGGGAGATGTTTATTT CCCCTGTTTCT	TTTTTGTGTGGATCAGCCCTAACA
466	rs3023377	12	113751107	CCCCTGTATCCTTTCCTCTTCTA	CCACAGCTCTTAAGTGTCCAAGT	CACGCACTTGCTTTTTGC
817	rs3664582	multiple		TGGGTAGGCTCACAGACTCC	GCCTATGAACCTTACCCTTTCA	TTTTTAAGAAGTGTCCCAAATCCTTTCTATATAGT
81	rs3023379	13	18303792	CCAGCCTAGGATGTGTTCAA	TACCCTTCACCACCACACAA	TTTTATATTTATCTCACTGTGAAGTCTGCCTA
892	rs3717068	13	30844331	AGTCCCACCCAAAGCATACC	GAAAGTCTGGCAATTCCTGTT	TTTTTTTTTTTTAGTCTACTTTCAGTGCTGTCCCAT
79	rs3089102	13	44412430	CCATGAGCCTTGAAGAAGGA	TGTACACATGAGACTCCGGC	TTTTTTTGAGTGTCACATGGATTTGCAC
484	rs3089436	13	60912145	AACTGACAGGGTGGTTGGTC	CATTTTGGAAATGTCGGCTT	CACATGGGGACTGTCCAAA
893	rs3688361	13	80402371	AACTGGAAAATCTAAATGAAATGG	TGATGGATACTGTCAAAGGTTCT	TTTTTTTCCTTAGCAGTTTAGGAATATTTAGATAGTTAA
80	rs3023386	13	96657145	ACCTCAGATAAGCGACATTTTCC	TATACATGCAGCCAAAACACCT	TTTTTTTTAAGTTCATCATTCCCTAGGATGTTATA
813	rs3144879	13	101262056	GAAGACCTCAGGGATCGTCA	ACTTTCCGCCACATGAACTT	GCCTCGGCACTGACCA
313	rs3023394	13	114216919	TAATACGACTCACTATAGGGAGATGATGTGAA ACCAGGAACCA	AATTAACCCTCACTAAAGGGAGACAGAGTTT CAGAGGGAGCTGA	TTTCAGATGCCGAGTAGCAGG
309	rs3023392	13	114837305	TTGATCTCACATCCATTCTTGG	AAAACCTAATGGTGTCAGAAGGC	TTTTAAGTAAATGAATGTGTTTATTGCTTCTC
814	rs3708958	13	117780654	CCTGGGGAGGGTGAAATTAT	AAGGCAGAGGCAAGAAAACA	GCTGAGTCACGGTACATAAAGTTGT
786	rs3720966	14	10727081	CAGAGGACCCATGTTGGAGT	GGTGTGCACAGAAACGAGAA	TTTTTTTTCACCCACGTTGTGACTTTAAGC
787	rs3712403	14	13327515	GCCTGACCAGGGTCATAAGT	TCAGATGTCTGGAGCCTGAA	CCTTCCTGTCTGTTCCAGC
788	rs3713871	14	24399404	CTGCCTGGTGCTAACCTCTC	AGAGAATGGGGGCCAATAAC	TTTTTTTTCCCTTAGCTTTCAAGTCCTTGC
809	rs3724533	14	27108284	AGGCTTCCCTGTGCACACTA	TTCCAAGCAATTACCCTGAGA	TTTTTTGGTAGGCAGGTACCAGAATCTCA
789	rs3713838	14	27644295	GAGGGATGGATGTAACCATGA	GGCACACCCAGCATAAGTTT	TTTTTTTTTTTGAGGGTCAGAGCACTTGCAGTA
790	rs3685393	14	30183496	GGCCAGGAGAGGTGATACTG	CCTGTTGCTTTTGCAGAGAC	CAGTTCCAAGCACCCACA
811	rs3023409	14	48748605	AAAATGGTTGGGCAAAAATG	GAGCTATGGTCCCAAATAGCC	TTGTGATATGTGGAAGTTATATAAGCTTC
89	rs3023408	14	50023498	TAGCCCAGAACACTGGTTCC	TTCCAGACTGACTCTTAATGGC	GCCCCTGACAATGAACG
321	rs3021908	14	53097465	TGAGAATGGTAGTTGTTATCCTCTG	TTTGAGACAGGGTGTCATGC	TTTTTTTTTTGACGGTGTGTGCCTACAAT
86	rs3089070	14	67715311	CTTGTGCACCTCAGCTATGC	GTGGGGGGACACAAAGACTTG	TTTCCTAGTTATTTCTGTGGTCTT
857	rs3705482	14	82835317	CACATTGTGCCAGAGAAATGA	TCAAACACCGGAAACTGAAA	TTTTTTAATATATATCCAGTGGAATTGAGTGGT
87	rs3088599	14	84734491	CCTCTGAATGCCTAAGCAGG	TGCTTGACACAAGGCTATGC	GCAACATATGAAGTAAATAATGTCTTAGC
858	rs3674616	14	98406482	GAACCTTAGGGATCCAGATGAA	GGCTTAACCAACGCTCTGTC	ATGCCCAGAGAGTGATCTAGAAG
88	rs3090773	14	105296240	TTCTTGCTCAGGGAATGGTC	AGCAATCACCTTGCTGGAAT	TTTTTTTTTAACTCACCTTTGTTTAGACAATCAGT
812	rs3657504	14	118523326	CTGGCTCAGGCTTTTGAAAC	GAATTGGAATGGCTCCCATA	TTTTTTTTTTTTTCAAGACCACATCTCCTATTCCTTCT
91	rs3023415	15	10765192	CACTGTTGCGAACTGCTCAT	GCACAGTCTGAGAACTCCCC	TCTCATTGGAACAAGTGAAAGC
894	rs3088491	15	12374865	TCCTGCTGTTTCCTGTAGGC	TTCCACATACAGTGCCCTCA	TTTTTTTTTTTCTTCAGACCCGTAGTCCTAAGGGCATG
895	rs3662097	15	20153137	TGTTTTATCACGCCTTCACC	TGCTACCCCAGTCCTCTTGT	TTTTTTTTTTTTATTGGACTCTATATTTTGTCTGGGCT
96	rs3023416	15	31860737	TTGCAGGTGTCTTTTATCTTCC	GACAGAGACTTTGGGGATGC	TTTTTCCCTTCAGAAATGAAAATTAATCTACTA
808	rs3669262	15	40601825	GCAAGGCCTAGACAGTATCCA	TTCCCTCATGGCTTGAAGAT	TTATCAGCACCCGTCCCA
92	rs3088488	15	42420068	CCAGGGCTTGAGTTTGGATA	CCCCACACACTGTGACAATC	GATCCTGAGGGATTGAACTC
859	rs3658370	15	51501719	TCAAGAGGAGGCTTTTGTGC	CTGTTTATGCGAACCAAGCA	TTTTTTTTTTCAATGTGGGTAAAACTGGCAAT
860	rs3701351	15	61455192	GCTGCCAAGTGAATGGATCT	TGCCCTCTCCAATATGATGA	TTTTTTTTGCCAAACAATTGTAAGAATGTGTG
93	rs3088506	15	71350770	GATTTGTCCCAGTTGTGCCT	CCTCACAGGGTCTCCTCTTG	TAAAACTCTCTCATGACCATAAATGTC
486	rs3090719	15	74403438	GGCGCTCTTTTTGAGAAATG	CCCCAATCCAGTCTGACAGT	TTTTGGTCCCACCTTGTTACAGGTC
94	rs3088710	15	86619624	GTGGCTAGTGCCTCTCACAA	TACTGGGTGAGTCTGTGGGC	TTACATTCATGTATACAGACCAGGAC
913	rs3677860	15	87544783	GCTCATATTTGTTGGCGACTG	CAAGTGACGCTGGTTTGTGT	TTTTTTTTTTCTGGCTGTTCATTATTTGTGCA
810	rs3717898	15	95702188	GGCTGTCCTGGAACTCACTC	CAGCTAGCCTTACCTCCCATT	ATTAAAGGCATGCACCACCA
101	rs3090912	16	6099904	GGATGACAAACAATGGCAGA	CAGATCCATCAGTGGTGGTG	TTTTTTTAAGTACTGATGGCTTTGAGTCTTA
806	rs3667072	16	16273453	AGGAGTGAAAAGCCAAGCAA	ATGGCATATGTTGCAGGTGA	TCATGATAGCCATGAGCACTC
332	rs3089488	16	18511345	GCTCTGAGCAGCTGGAGTTT	ACGTGCACACACACCTCAGT	AGTAAATAAGTTTGAACACTCTAAAAGT
880	rs3695744	16	27296228	ACGTGTTCTTATCTTGTGGAGAC	GCTATGGAGGTTTTGAAGTATGG	TTTTTTTTTTTCACCTTGCTAGGTTTTTATTGGTAC
881	rs3718034	16	35665619	CTCTCTCTCACTTTCTGTTTGTTTG	TGCCTTTAATCCCAGCACTC	AGGCAGATCTCTGAGTTCGAGG
487	rs3089787	16	38850538	GGGGTCTGTTCGTCTCAAAA	ATAGAGTCAAAGGGGTGGGG	AATGTATTTTTTAATTGGTTATTTTACTTATTAA
488	rs3090260	16	45609646	TCCCCTTAAAGTCTCAACTCCAT	CCTGCCTAAATCAGTAGGCATAA	TTAAAGTCTCAACTCCATCTTTCC
794	rs3654982	16	63555294	TTAATCTGGCCTCTTGAGCTG	AGCGGATTGAAAACATCTGC	TCTAGTCCACCAGCAGCAGAAAC
795	rs3719654	16	70347393	AAGGTCAGAGCATGCCAAGT	CCACACCTCCTAAACCTACTCC	CGTCTCATCAAATTGCTATTCCA
489	rs3090908	16	72217393	TCCAAAATCAGGAGGGAAAA	GCCTCAAAACGAGCACAGAC	CAGAAACATAATTTCAAAGTTGCA
796	rs3695101	16	76223854	GGGAGGCACTACACACACCT	CAGGCTCAGCATCAGCAGTA	ATTCAGGGAGAAGAGGACACCT
490	rs3090645	16	85879752	GCAACAAGTATTTGAAGATGAAACA	TTGCAGAATACCAGAGGGCT	TTTTTTGGCTACTACACTAGCAAATCCATAG
455	rs3023436	16	88046993	GGGCATCATCTATTGGGTTG	GAATGCCTCTTTGACAATGGA	TTTTCCAGGGAGTAAAACATCAGG
304	rs3023441	16	97552996	TAATACGACTCACTATAGGGAGAAAGCAAGC AAGCCCTTTGTA	AATTAACCCTCACTAAAGGGAGAATCCTGAG AGGTGCCAT	TTTTTTTTTAGAGTCATTCTACTTCAGCAGGAAG
882	rs3674239	17	12357534	GCCATCAAGGACTTTTCATTTTC	GGGTCCATATCCCACTGCT	TCTTCCACTCTTTGGTCCTCC
883	rs3668662	multiple		AAATACCGGCCCATCTCTCT	GTGCTAACTGGTGCCTGAAA	AAAGGCAGTGGGTACACATCAT

104	rs3090500	17	25071190	ATCCCTGCCTTCCTCTCCT	CGCATTTTATTGAGGCATAGC	TTTTCCTTCCACTCAAGTAGCTACACTG
884	rs3708501	17	30648071	CACCTGGCCTTCAGTGTTTG	TGCTAAATGTGGGTGTGGTATG	AGGTGGCCGCTGCCA
491	rs3088914	17	36695497	CCATGCTGTCCTCAAACTCA	GCAGAAAGAGGGGAACTGTG	GTTTCATTCCTTCAAGTTCTGAA
105	rs3023454	17	36882223	TAATACGACTCACTATAGGGAGATCCCTGAG GACTGAGTCTCCT	AATTAACCCTCACTAAAGGGAGAGGAAGAA GCTGCCCATAGAA	TTTTCCAGGACTTCTTCATCCC
452	rs3023449	17	42120599	AATGAACAACGGGCAGGTAG	GCCCCAGGATGTCATTATTT	TGACACAACGCTGTTCCAGG
862	rs3700023	17	52981854	AGACCTTGACTTTACCACACACA	TCTGTGACCTGGCAGATGAT	TTTTTTTTGATTGCCCTATAGCCATTACCTG
453	rs3023456	17	67310979	CCAGCTCCAATCAGTCCAAT	GAAGCACCCAGAAATCAGGA	TTTTTTTTTATGTGCCTTCTTATGTCGACTTTAT
863	rs3712928	17	78656292	CGGAAGATGGAAAGGAGTCA	TCGAAAGATTGATGGTGGAA	GGAGCTCCCATGGAGAGG
885	rs3710028	17	83573655	CGCACAAGAGCCTGGATT	CTGCCTCTGCCTCCCAAGT	TTTTTTTTTTTGGCTTATTTTTTCTTGTTTGAGACAG
312	rs3089323	17	83844781	TGGCCTTGTGGAAGAGATGT	AGAGCACTTGGAAGGCTGAG	TTTTTACAAGCTAGTCTCAAACTTGTAGCA
864	rs3668190	17	91677947	GATTAAAGGTGAGGTTCACCAAA	GGAGGCAATGACAAGCAGAT	TTTAGAACTTGCTCTGTTCTAGGCTG
886	rs3696042	18	8912956	GACTTCACAATAAAAGATGCACA	AGGACAGTGGCATAGGAAGA	TTTTTTTCTAAACTCAGTTTCTATTGATTAAATGCTTC
865	rs3655356	18	10241322	TGCTGAAATCCCCCAGTATC	CCAATCACCATTCTCCCCTA	ACGAGCAAATGCTGAATCTCA
887	rs3695261	18	21869777	TTCCTTCTGGTTGGAATTGT	CACATGCTCTGTAGCTCCTTTT	TTTTTTTTTGGAATTGTATTAGATCAATGGTCCA
450	rs3023463	18	29943155	CCTCTGGACACATTCTTCTCG	CCCAGTCTCCACTGTAAAGCA	TTGCTCAGACTATCAGTGCTACTTCTAGAA
803	rs3683689	18	33935868	GCATGCGTGTACTTTGCAGT	GCACTCATCGGTCTTTCTCC	TTTTTTCACTGTGGTCAGACAGAATGCA
316	rs3089327	18	41908447	GTTCCTGGTGGTAGGGGTTT	ACTGAGGAGCAGAAAAGGCA	TTTTTTTTTTTTTTTTTGATGGCTTCTGCTTTCCTTT
888	rs3657200	18	50109691	AATCTCCATCCACCAAAAGG	TCACCCATGAAACAAACACA	TTTTTTTTAATTCGCCAATGCAGGAGTACC
109	rs3023468	18	63609445	TTCGGTTACCACACACCATT	AGCCATGCTGGTATTTCCTG	GCACATGGTTCTGCCACA
866	rs3657018	18	78455291	TGAAGAGTCAAAACCCACAGC	GGAACACAATGGGTGGAATC	TTTTAGTTCTTAAGTACCCTATTCTGAGCA
889	rs3668347	18	80932716	GCTTGTCCTGATGCAGACTACTAA	TTCAAATGCCAGGTGTCAAG	GCACTCCGTTTGAGAACACTC
804	rs3718427	18	88768623	CCCTGGGGCATCAAGTCTAT	GTTGTGTGGAGCATCGTTTG	AAAGAACATAAATGGGCTATTCTTAG
867	rs3665935	18	89757219	CCATTAATCAGCAGGCTCGT	TGAGGATCACTCTAAACACAGGA	TTTTCTTGCTCCCATTATGATGCA
800	rs3671678	19	17422426	TACCCCTCCTCTTCCTTTCC	CTGGTGCGTTGACTTGAGAA	TTTTTTTTTGTGGTGGGGGAGCAATATG
891	rs3669192	19	16913884	GATGCTATGGCTGCTTTGCT	CCATTTTCTCTTAACTTTGGTTTC	TTTTTTTTGTGTGTGTTTGTTTGAGGTTGA
890	rs3658201	19	20407978	AGGTATGTTGAAGGTATGGATTATTT	CGTGCGGAAGGAAGTTGAG	TTTTTTTTTTTTGGTTTTGATTCTTCAGTGTAGTTTGG
111	rs3023481	19	20484928	AGGGCAGAACAACAAGGAGA	GCACTTTCCACAGTTGGGTT	ACAAGGAGACTATATATCTACTTTATGTGGA
868	rs3710059	19	30660820	AGTTGATTTGCCCTGTGGAG	CATTGTGGACACCAGACAGC	TTTTGGAGCTGGCACATCCACT
336	rs3090951	19	44127717	AGATGGCTGTGATCGGAAAC	CCCTCATTAGTTGCGGTGTT	TTTTTTTAGTAGCTTGGTGTTTTGTTTGTTT
801	rs3726430	19	44193090	GCCCTACTGGCAGCCTTACT	GGGGTTTGTGAGAGCTGTGT	GGGTAGGGGTAGGAAGTAGAGAG
869	rs3713040	19	55813052	TGTTAAGGCAGTGTGTCGTG	AAGCAAGCGCTAACACCAGA	TGAAGCAGGAGCACTGAGC
449	rs3023498	19	60821445	CCTAGACGAGGGATGGATGA	CTGCACCTCTCCTGGAAATC	TTTTTTTTTGGGCTGGATGATGACCG
802	rs3685993	19	60910395	CTCAAACGTGGGGGACATT	GATCCATCTTTCTCTGCCTGT	GAAATGGCTCATTCAAGTAAAGTG
492	rs3088634	15	19069275	CCCTAAACCCTAGCCCTGAC	GAAATTGATGGGTGGCATCT	GGGGAAGGTGGAAGTGAGTATAT
605	rs3089544	18	5088109	CATGGTCACCTGGCAACA	TCTGGTGCAATTCTACAACCC	TTTTTGACAGACAGCAGGTCAACGA
604	rs3088804	x	99074052	GATGGCTAATTCCAGGCCTAC	CAACAGCAGAGCTGTACCCA	CAAGCTTCATTTAGAAATACTTTGC

Supplemental Table 3: PCR multiplexing strategy in genome-wide screen

Primers are same as listed in Supplemental Table 1

SBE group	PCR grp	Assay ID	RefSNP ID
1A	1a-1	412	rs3089257
		405	rs3022887
	1a-2	772	rs3704164
		426	rs3090833
	1a-3	471	rs3022975
		779	rs3660209
	1a-4	786	rs3720966
		830	rs3714631
	1a-5	883	rs3668662
		850	rs3090435
	1a-6	439	rs3023183
1B	1b-1	452	rs3023449
		823	rs3713224
	1b-2	109	rs3023468
		868	rs3710059
	1b-3	893	rs3688361
	1b-4	455	rs3023436
	1b-5	879	rs3695889
	1b-6	490	rs3090645
	1b-7	922	unmapped
1C	1c-1	461	rs3023258
		453	rs3023456
	1c-2	821	rs3704980
		600	rs3022802
	1c-3	808	rs3669262
		895	rs3662097
	1c-4	415	rs3022979
		840	rs3694308
	1c-5	422	rs3023045
	1c-6	79	rs3089102
1D	1d-1	33	rs3023057
		713	rs3684370
	1d-2	761	rs3665023
		489	rs3090908
	1d-3	803	rs3683689
		831	rs3676476
	1d-4	81	rs3023379
		892	rs3717068
	1d-5	757	rs3682376
	1d-6	478	rs3022989
1E	1e-1	741	rs3696551
		700	rs3659426
	1e-2	770	rs4137954
		756	rs3688884
	1e-3	790	rs3685393
		51	rs3023194
	1e-4	762	rs3679837
		701	rs3681957
	1e-5	911	rs3697014
		812	rs3657504
	1e-6	913	rs3677860
1F	1f-1	797	rs3696966
		839	rs3667625
	1f-2	727	rs3681675
		857	rs3705482
	1f-3	746	rs3657668
		809	rs3724533
	1f-4	882	rs3674239
		768	rs3716232
	1f-5	769	rs3694785
	1f-6	898	rs3667466
1G	1g-1	833	rs3726717
		843	rs3657720
	1g-2	844	rs3691937
		834	rs3689513
	1g-3	827	rs3090608
		817	rs3664582
	1g-4	912	rs3702150
	1g-5	763	rs3686956
	1g-6	742	rs3722968
	1g-7	814	rs3708958
1H	1h-1	766	rs3724779
		465	rs3088800
	1h-2	788	rs3713871
		436	rs3089474
	1h-3	856	rs3697769
		75	rs3023347
	1h-4	715	rs3704392
		835	rs3669413
	1h-5	86	rs3089070

group	PCR grp	Assay ID	RefSNP ID
2A	2a-1	4	rs3022839
		96	rs3023416
	2a-2	446	rs3090586
		488	rs3090260
	2a-3	32	rs3023037
		445	rs3089912
	2a-4	721	rs3664018
		859	rs3658370
	2a-5	218	rs3023026
	2a-6	820	rs3708255
	2a-7	80	rs3023386
2B	2b-1	818	rs3722942
		822	rs3685188
	2b-2	722	rs3681847
		732	rs3706063
	2b-3	339	rs3023161
		811	rs3023409
	2b-4	858	rs3674616
		118	rs3023175
	2b-5	764	rs3711535
2C	2c-1	20	rs3022960
		483	rs3088822
	2c-2	448	rs3023243
		860	rs3701351
	2c-3	67	rs3023265
		321	rs3021908
	2c-4	724	rs3664805
		826	rs3663534
	2c-5	800	rs3671678
	2c-6	201	rs3023256
	2c-7	890	rs3658201
2D	2d-1	854	rs3710192
		851	rs3723894
	2d-2	787	rs3712403
	2d-3	794	rs3654982
	2d-4	845	rs3023067
2E	2e-1	726	rs3667376
		731	rs3674631
	2e-2	767	rs3672332
		435	rs3090731
	2e-3	862	rs3700023
		832	rs3669022
	2e-4	606	rs3023117
		829	rs3660910
	2e-5	750	rs3709317
	2e-6	740	rs3666032
	2e-7	738	rs3711350
2F	2f-1	486	rs3090719
		745	rs3713298
	2f-2	801	rs3726430
		101	rs3090912
	2f-3	484	rs3089436
		733	rs3672323
	2f-4	841	rs3707288
		789	rs3713838
	2f-5	909	rs3721297
2G	2g-1	714	rs3662163
		816	rs3726591
	2g-2	463	rs3088501
		723	rs4137557
	2g-3	413	rs3090381
		799	rs3719410
	2g-4	37	rs3023051
	2g-5	730	rs3666331
	2g-6	748	rs3685067
	2g-7	828	rs3706262
	2g-8	900	rs3716435

Chapter 4

Final Discussion

The sequencing of the human and mouse genomes has spurred a growing interest in analyzing mouse models of human cancer using genomic techniques. Comparative genomic studies on mouse vs. human tumors can be valuable in two major ways: 1) in validating mouse models through an assessment of their degree of genetic resemblance to human disease and 2) in identifying genes and/or gene sets that are common to mouse and human tumorigenesis. As described in Chapter 1, many analytic tools have emerged in recent years for human genome mining. Some of these tools have been translated to murine versions. The work described in Chapters 2 and 3 involved application testing of two newly translated mouse whole-genome analytic techniques: ROMA and SNaPshot (Applied Biosystems) SNP genotyping. With ROMA, a high-resolution view of copy number alterations in the tumor genome was possible. By SNaPshot (Applied Biosystems), low-density SNP-based draft maps indicative of LOH were obtainable. The applications were tested mainly on a murine model of lung cancer. Several recurrent chromosomal copy number gains and losses, as well as chromosomal LOH, were observed in this *Kras*-driven lung cancer model. Discussed in this concluding chapter are my views on the technology applied and the biology implied by the body of work.

On Technology

A wide range of copy number alterations in primary mouse tumors has been previously documented, including single copy gain or loss of entire chromosomes, partial gain or loss of a chromosome, high-amplitude focal amplifications, to low-level small deletions (Hodgson et al. 2001; You et al. 2002; Hackett et al. 2003; O'Hagan et al. 2003). To cover this broad spectrum, a genome-wide high-resolution CGH tool would be invaluable

cover this broad spectrum, a genome-wide high-resolution CGH tool would be invaluable for mouse cancer DNA anlayses. Among the many CGH platforms for genome-wide copy number analysis studies in human, ROMA has one of the highest resolving power averaging at 30kb (Lucito et al. 2003). Tested in Chapter 2, the mouse version of ROMA also appeared to be a powerful tool, allowing us to detect a focal high-amplitude (>4.6fold, assessed by Southern) *N-Myc* amplification in retinoblastomas of a *Rb/p130* DKO model, as well as numerous whole-chromosomal gains and losses in the same retinoblastoma sample set and in the lung tumors driven by a *Kras* mutation.

The mouse ROMA platform would be a useful tool to characterize mouse cancer-associated genetic alterations. Genome-wide copy number data from mouse tumors may potentially be used in several ways: 1) clear focal changes can pinpoint individual candidates such as *N-Myc* in our case; 2) clustering of data can identify copy number alteration patterns that define particular tumor subtypes in the mouse (O'Hagan et al. 2003); 3) analogous to the use of gene-set enrichment analysis (GSEA) approach to extract human tumor gene-expression signature from mouse tumor gene-expression pattern (Sweet-Cordero et al. 2005), one can imagine being able to extract human copy number change signatures by comparing with mouse data that are confounded with less genetic noise.

There is also an abundant choice of SNP genotyping methodologies, although few have been applied for whole-genome analysis in the mouse. We attempted to contribute the following ways: 1) validate 358 published SNPs individually for polymorphism in

129S4/svJae vs. C57BL/6J strains, 2) worked out a protocol to genotype a panel of 147 markers in the mouse using SNaPshot (Applied Biosystems) genotyping system. The method uses a standard DNA sequencing machine to resolve SBE products of SNPs, which should be easily adaptable for use by interested labs, 3) tested the application of the assay as a mapping tool of LOH in mouse tumors. Although genome-wide SNP mapping of LOH is not new in human cancer genomics, the same concept has not been performed in mice probably due to the lack of accessible SNP genotyping protocols.

Despite needing further improvements on overall genotyping accuracy and efficiency, our protocol was usable in analyzing LOH patterns in mouse lung tumors. We correctly detected the loss of wild-type p53 allele in a subset of samples, suggesting the concept of performing SNP-based LOH detection in F1 mice is going to be viable. When other groups concentrated their efforts in the genotyping a few markers in multiplex with the same type of assay, they were able to drive the level accuracy higher (to almost 100%) (Makridakis and Reichardt 2001; Norton et al. 2002). This suggests there is much room for us to optimize each of our SBE reaction to improve on accuracy. In addition, sensitivity of the method needs to be tested on more markers in a dilution experiment using heterozygous vs. different homozygous DNA. Then, in the future, increasing the density of SNPs in the panel would allow higher resolution.

To date, over 6 million mouse SNPs across different strains have been referenced in the NCBI dbSNP public database. The challenge in the future SNP-based LOH mapping clearly is not the lack of markers, but is finding an efficient and cost-effective method to

genotype SNPs at a density an experiment requires, which may change as the project proceeds. Flexibility in genotyping the particular SNPs of choice will also be required. SNaPshot is a multiplexable system, which we used to genotype 5-11 SNPs in each SBE reaction and capillary run. No investment in specialized instrument is required. However, I think the biggest charm of the system is the flexibility in choosing the markers used in genotyping. The assay can be performed either individually or through mixing-and-matching and adding-and-removing of a few markers in a multiplex fashion.

On Biology

The applications of ROMA and SNaPshot genotyping were tested mainly on murine models of lung cancer that were initiated by a *Kras* mutation. Overall, copy number alterations or LOH seem to start appearing in genomes of tumors that are histologically graded 2-3, when they started to exhibit pleomorphic nuclei, prominent nucleoli, and nuclear molding. The correlation of detectable genetic alterations with higher histological grades fits the clonal evolution model of tumor progression. However, lower grade tumors, because of their smaller sizes, are more difficult to dissect out cleanly. We attempted to minimize this confounding factor by choosing samples with little normal tissues, as judged histologically. Laser capture microdissection could be used in the future as another way to circumvent this contamination possibility.

Based on histopathology, lung cancer is grouped into two broad categories: non-small-cell lung cancer (NSCLC) and small-cell lung cancer (SCLC). About 80% of all cases are NSCLC, which are further divided into the following subtypes: adenocarcinoma, squamous cell carcinoma, bronchioalveolar carcinoma, and large-cell carcinoma. The

NSCLC subtypes behave similarly as a distinct group to SCLC in its therapeutic response (Minna et al. 2002).

Genetically, SCLC and NSCLC also form two distinct groups. Genetic alterations manifest as large-scale chromosomal gains and losses, focal amplifications and deletions, or nucleotide changes. Multiple mutations are observed in lung cancer samples. Epigenetic changes such as methylation also occur commonly. Among the oncogenes, *Ras* mutations are detected in 20-50% of NSCLC and <1 % of SCLC (Slebos et al. 1990). *Kras* mutations consist of 90% of all *Ras* mutations and are almost exclusively found in adenocarcinomas (Slebos et al. 1990). *Myc* amplification or overexpression is seen more frequently in SCLC (20-35%) than in NSCLC (5-20%) (Richardson and Johnson 1993). Bcl2 is overexpressed in 75-95% of SCLC and 10-30% of NSCLC (Pezzella et al. 1993; Kaiser et al. 1996). *EGFR* mutation is found in 20% of NSCLC and is associated with non-smoking related adenocarcinomas (Zochbauer-Muller et al. 2002). Among the tumor suppressor genes, *p53* is found deleted or mutated in >50% of NSCLC and SCLC (Takahashi et al. 1989; Toyooka et al. 2003). $p16^{INK4a}$ hypermethylation or deletion occurs frequently; ~30-50% NSCLC does not express *p16* (Minna et al. 2002). The alternative reading frame product in the same locus $p14^{ARF}$ is also inactivated in ~20% of NSCLC and ~65% of SCLC (Nicholson et al. 2001). In addition to gene-specific lesions, cytogenetic, CGH, and LOH studies have revealed numerous large-scale chromosomal aberrations, suggesting more oncogenes and tumor suppressors remain to be discovered. LOH in chromosome 3p is the most prominent event, found in almost 100% of SCLC and >90% of NSCLC (Wistuba et al. 2001; Zabarovsky et al. 2002). Other frequent changes

include gains of 1q, 3q, 5p, 8q, 11q, 12q, 19q, and losses of 4q, 10q in NSCLC, and gains of 3q, 5p, 8q, 19q and losses of 4q, 5q, and 13q in SCLC (Balsara and Testa 2002).

The tumors developed in the *Kras* model histologically resemble human NSCLC (Jackson et al. 2005). Recurrent chromosomal copy number changes include +6, +12, +19, + 3, + 16, -9, -11 in 10/28 lung tumors. Each of these chromosomes is comprised of multiple syntenic human chromosomal regions. Genetic alterations in many of these regions have been observed in human lung cancers as summarized in Table 1. Chromosome 6 gain was seen in 80% of tumors we tested. Mouse chromosome 6 harbors the *Kras* gene and contains a region in synteny to a human 3q segment that have been implicated in lung cancer. Also of interest are the recurrent losses of chromosomes 9 and 11, with each reduced in copy in two tumors. A distal part of chromosome 9 is syntenic to human chromosome 3p21-22. Loss of human chromosome 3p is the most common event observed in lung cancer (Zabarovsky et al. 2002). In particular 3p21 loss is observed as an early event, which can be detected in the pre-malignant epithelium of smokers (Zabarovsky et al. 2002). As for chromosome 11, its distal arm has syntenic conservation to the entire human chromosome 17, where *p53* resides. It is reassuring that the lung tumors developed in our mouse model contain certain regions that are syntenic to the regions altered in some human lung cancers. However, the genetic lesions observed in our mouse model span entire chromosomes, thus the subchromosomal regions critical to tumorigenesis in our model are unclear.

Table 1: Summary of copy number changes observed in mouse mouse model and their corresponding syntenic regions in human

Recurrent changes in mouse model	Human synteny	Reported gains in human NSCLC	Reported losses or LOH in human NSCLC	Potential genes
+6	7q		Sasotomi (2002), Wong (2002)	
	7p	Testa (1994)		
	4q			
	2p			
	2q			
	3q	Pei (2001)		
	3p		Sasotomi (2002), Testa (1994), Zhao(2005), Yoshimo (2003)	
	12p			Kras2
+12	2p			
	7p	Testa (1994)		
	7q		Sasotomi (2002)	
	14q			
+19	11q	Testa (1994)	Pan (2005)	
	9p		Zhao (2005), Pan (2005), Yoshimo (2003), Girard (2000)	
	9q		Testa (1994)	
	10q		Zhao (2005)	
+3	8q	Pei (2001)	Girard (2000)	
	3q			
	4q			
	1q	Pei (2001), Testa (1994)		
	1p		Pan (2005)	
+16	16p			
	8q	Pei (2001)		
	22q			
	3q	Pei (2001)	Zhao (2005), Pan (2005)	
	3p		Sasotomi (2002), Testa (1994), Zhao(2005)	
	21q		Testa (1994)	
-9	11q	Testa (1994)		
	19p		Testa (1994)	
	15q			
	6p			
	6q		Testa (1994), Girard (2000)	
	3q	Pei (2001)	Zhao (2005), Pan (2005)	
	3p		Sasotomi (2002), Testa (1994), Zhao(2005)	RassF1
-11	22q	Zhao (2005)	Testa (1994)	
	7p	Testa (1994)	Sasotomi (2002)	
	2p		Zhao (2005)	
	16p			
	5q		Sasotomi (2002), Pan (2005)	
	17p		Testa (1994), Pan (2005), Wong (2002)	
	17q		Wong (2002)	

Our model is driven by a *Kras* point mutation. In human, *Kras* point mutations occur at a higher frequency in smokers than in nonsmokers; one study reported the numbers to be 30% vs. 7% (Westra et al. 1993). Smoking is the biggest risk factor for lung cancer. Differences in the spectra of genetic alterations in smoking vs. non-smoking related lung cancer have been observed (Hirao et al. 2001; Sanchez-Cespedes et al. 2001; Wong et al. 2002; Sy et al. 2003; Wong et al. 2003). Cigarette smoke consists of multiple carcinogens including benzo(a)pyrene (BAP) and 4-methylnitrosamino-1-3-pyridal-1-butanone (NNK), which result in a prevalence of G-to-T and G-A transversions in smoking-related lung cancer (DeMarini 2004). The spectra of mutations differ in smoking vs. non-smoking related lung cancers (Pan et al. 2005). Interestingly, 1p, 3p, 5q, 11q, and 17q, which share homology with gained or lost in the DNA from our *Kras*-induced mouse model, have been reported in the cited studies.

Despite a certain degree of syntenic conservation of lung cancer genetic changes could be inferred, one apparent difference is that the predominance of whole-chromosomal changes in the lung tumors from mice. Such observation seems to be common among spontaneous tumors of different genetically engineered mouse models. On the other hand, multiple number and kinds of mutations, including translocations and sub-chromosomal lesions, are frequently seen in human carcinomas. At the cellular level, mouse cells differ from human cells by having longer telomeres, which might have protected mouse chromosomes from breakage events through break-fusion-bridge cycles caused by damaged telomere (O'Hagan et al. 2002). Indeed, tumors in *mTerc*-/- mice with deficient telomerase had a larger and more human-like variety of genetic changes (O'Hagan et al.

2002). Perhaps making mice containing both mTerc deletion and Kras mutation in the lung will lead more human-like focal genetic lesions in the lung tumors.

Another point of interest is The set of recurrent chromosomal copy number changes we observed (in order of prevelance: +6, +12, +19, + 3, + 16, -9, -11) in 10/28 lung tumors by ROMA were different from a previously reported set of 16/59 samples from a closely related lung cancer model (+6, +8, +16, +19, -4, -11, -17) analyzed using BAC array CGH (Sweet-Cordero et al. 2006). Chromosome 6 and 19 gains and 11 losses are the only commonalities. While chromosome 12 gain was seen in 5/10 tumors with changes in our study, it was observed in 1 clear case in the other analysis.

The discrepancy between the 2 studies raises a few questions. The discrepancy could be simply due to small sampling sizes and the random nature of mutational process. Formally, a pilot study can be initially performed to generate a statistical estimate of the amount of samples required in the main study. However, in practice, this is not often done, largely due to cost constraints and in the case of a human study, also because of the difficulty in obtaining patient samples. Meta-analyses may help to make sense of studies done in different times or labs (Hoglund et al. 2004). Alternatively, this discrepancy may reflect a true difference due to the different activation timing/mechanism between the conditional $Kras^{LSLG12D}$ and latent $Kras^{G12DLA2}$ alleles. Our study was performed using tumors materials obtained from $Kras^{LSLG12D}$;*p53*+/+, $Kras^{LSLG12D}$; *p53*+/- and $Kras^{LSLG12D}$; $p53^{R270H}$/+ mice while the other study used mice containing a $Kras^{G12DLA2}$ allele. However, in our study, tumors from mice of all 3 genotypes that were comparable

in histology were also comparable in the types and amount of genetic changes. The two *Kras*^{G12D} alleles lead to lung tumors with indistinguishable histology, despite the different timing/mechanism of activation. This would be reminiscent of the RIP-Rag islet cell cancer model, in which the timing of T antigen activation changes the pattern of copy number alterations without affecting tumor histology/progression (Hager et al. 2004). In addition, genetic background may contribute to the discrepancy by affecting copy number changes. One difference in the two experimental setups was the background of mice used: our analysis was on an inbred 129S4 strain while the other was done on F1 B6x129S4 tumors. A difference in tumor incidence in B6 vs. 129S4 strains has been observed in Kras^{G12DLA2} mice, suggesting there are strain-specific modifying factors for tumor multiplicity. Genetic background effect on copy number spectrum has been reported in another mouse cancer model (Hager et al. 2004).

The work of this thesis began the use of two powerful techniques: ROMA and SNP genotyping, to study mouse tumors. Our analysis of whole tumors has revealed a few stable DNA changes that might harvest critical genes for tumorigenesis. In the case of retinoblastoma, the candidate region was narrowed down to a single gene --*N-Myc*. The functional role of *N-Myc* in retinoblastoma remains to be characterized. In the case of lung cancer, an analysis of higher grade tumors will be worthwhile to attempt. Among the many unanswered questions surrounding the genetic alterations in cancer, whether genomic instability occurs is one of them. Genomic instability is a highly debated concept that was proposed to explain the origin of the many genetic alterations often seen in a cancer. Genomic instability describes an increased rate of genetic alterations in

cancer cells. Most of the described experiments were performed on DNA extracted from primary tumors. Such analyses in fact reflect the *stable* genetic changes that have undergone selection in the particular tumor. Measurement of genomic instability requires knowing the amount of changes that occur over a known number of cell division. Metaphase analysis of dividing tumor cells is probably the best measurement that can be done in tumors samples and can be included in future studies. Genomic analysis of tumorigenesis provides insight into the evolution of tumor cells. With the many new technologies, the coming years in cancer genome analytics will likely be full of excitements.

References

Balsara, B.R. and Testa, J.R. 2002. Chromosomal imbalances in human lung cancer. *Oncogene* **21**(45): 6877-6883.

DeMarini, D.M. 2004. Genotoxicity of tobacco smoke and tobacco smoke condensate: a review. *Mutat Res* **567**(2-3): 447-474.

Hackett, C.S., Hodgson, J.G., Law, M.E., Fridlyand, J., Osoegawa, K., de Jong, P.J., Nowak, N.J., Pinkel, D., Albertson, D.G., Jain, A., Jenkins, R., Gray, J.W., and Weiss, W.A. 2003. Genome-wide array CGH analysis of murine neuroblastoma reveals distinct genomic aberrations which parallel those in human tumors. *Cancer Res* **63**(17): 5266-5273.

Hager, J.H., Hodgson, J.G., Fridlyand, J., Hariono, S., Gray, J.W., and Hanahan, D. 2004. Oncogene expression and genetic background influence the frequency of DNA copy number abnormalities in mouse pancreatic islet cell carcinomas. *Cancer Res* **64**(7): 2406-2410.

Hirao, T., Nelson, H.H., Ashok, T.D., Wain, J.C., Mark, E.J., Christiani, D.C., Wiencke, J.K., and Kelsey, K.T. 2001. Tobacco smoke-induced DNA damage and an early age of smoking initiation induce chromosome loss at 3p21 in lung cancer. *Cancer Res* **61**(2): 612-615.

Hodgson, G., Hager, J.H., Volik, S., Hariono, S., Wernick, M., Moore, D., Nowak, N., 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**(4): 459-464.

Hoglund, M., Gisselsson, D., Hansen, G.B., and Mitelman, F. 2004. Statistical dissection of cytogenetic patterns in lung cancer reveals multiple modes of karyotypic evolution independent of histological classification. *Cancer Genet Cytogenet* **154**(2): 99-109.

Jackson, E.L., Olive, K.P., Tuveson, D.A., Bronson, R., Crowley, D., Brown, M., and Jacks, T. 2005. The differential effects of mutant p53 alleles on advanced murine lung cancer. *Cancer Res* **65**(22): 10280-10288.

Kaiser, U., Schilli, M., Haag, U., Neumann, K., Kreipe, H., Kogan, E., and Havemann, K. 1996. Expression of bcl-2--protein in small cell lung cancer. *Lung Cancer* **15**(1): 31-40.

Lucito, R., Healy, J., Alexander, J., Reiner, A., Esposito, D., Chi, M., Rodgers, L., Brady, A., Sebat, J., Troge, J., West, J.A., Rostan, S., Nguyen, K.C., Powers, S., Ye, K.Q., Olshen, A., Venkatraman, E., Norton, L., and Wigler, M. 2003. Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation. *Genome Res* **13**(10): 2291-2305.

Makridakis, N.M. and Reichardt, J.K. 2001. Multiplex automated primer extension analysis: simultaneous genotyping of several polymorphisms. *Biotechniques* **31**(6): 1374-1380.

Minna, J.D., Fong, K., Zochbauer-Muller, S., and Gazdar, A.F. 2002. Molecular pathogenesis of lung cancer and potential translational applications. *Cancer J* **8 Suppl 1**: S41-46.

Nicholson, S.A., Okby, N.T., Khan, M.A., Welsh, J.A., McMenamin, M.G., Travis, W.D., Jett, J.R., Tazelaar, H.D., Trastek, V., Pairolero, P.C., Corn, P.G., Herman, J.G., Liotta, L.A., Caporaso, N.E., and Harris, C.C. 2001. Alterations of p14ARF, p53, and p73 genes involved in the E2F-1-mediated apoptotic pathways in non-small cell lung carcinoma. *Cancer Res* **61**(14): 5636-5643.

Norton, N., Williams, N.M., Williams, H.J., Spurlock, G., Kirov, G., Morris, D.W., Hoogendoorn, B., Owen, M.J., and O'Donovan, M.C. 2002. Universal, robust,

highly quantitative SNP allele frequency measurement in DNA pools. *Hum Genet* **110**(5): 471-478.

O'Hagan, R.C., Brennan, C.W., Strahs, A., Zhang, X., Kannan, K., Donovan, M., Cauwels, C., Sharpless, N.E., Wong, W.H., and Chin, L. 2003. Array comparative genome hybridization for tumor classification and gene discovery in mouse models of malignant melanoma. *Cancer Res* **63**(17): 5352-5356.

O'Hagan, R.C., Chang, S., Maser, R.S., Mohan, R., Artandi, S.E., Chin, L., and DePinho, R.A. 2002. Telomere dysfunction provokes regional amplification and deletion in cancer genomes. *Cancer Cell* **2**(2): 149-155.

Pan, H., Califano, J., Ponte, J.F., Russo, A.L., Cheng, K.H., Thiagalingam, A., Nemani, P., Sidransky, D., and Thiagalingam, S. 2005. Loss of heterozygosity patterns provide fingerprints for genetic heterogeneity in multistep cancer progression of tobacco smoke-induced non-small cell lung cancer. *Cancer Res* **65**(5): 1664-1669.

Pezzella, F., Turley, H., Kuzu, I., Tungekar, M.F., Dunnill, M.S., Pierce, C.B., Harris, A., Gatter, K.C., and Mason, D.Y. 1993. bcl-2 protein in non-small-cell lung carcinoma. *N Engl J Med* **329**(10): 690-694.

Richardson, G.E. and Johnson, B.E. 1993. The biology of lung cancer. *Semin Oncol* **20**(2): 105-127.

Sanchez-Cespedes, M., Ahrendt, S.A., Piantadosi, S., Rosell, R., Monzo, M., Wu, L., Westra, W.H., Yang, S.C., Jen, J., and Sidransky, D. 2001. Chromosomal alterations in lung adenocarcinoma from smokers and nonsmokers. *Cancer Res* **61**(4): 1309-1313.

Slebos, R.J., Kibbelaar, R.E., Dalesio, O., Kooistra, A., Stam, J., Meijer, C.J., Wagenaar, S.S., Vanderschueren, R.G., van Zandwijk, N., Mooi, W.J., and et al. 1990. K-ras oncogene activation as a prognostic marker in adenocarcinoma of the lung. *N Engl J Med* **323**(9): 561-565.

Sweet-Cordero, A., Mukherjee, S., Subramanian, A., You, H., Roix, J.J., Ladd-Acosta, C., Mesirov, J., Golub, T.R., and Jacks, T. 2005. An oncogenic KRAS2 expression signature identified by cross-species gene-expression analysis. *Nat Genet* **37**(1): 48-55.

Sweet-Cordero, A., Tseng, G.C., You, H., Douglass, M., Huey, B., Albertson, D., and Jacks, T. 2006. Comparison of gene expression and DNA copy number changes in a murine model of lung cancer. *Genes Chromosomes Cancer* **45**(4): 338-348.

Sy, S.M., Wong, N., Mok, T.S., Tsao, M.S., Lee, T.W., Tse, G., Blackhall, F.H., Johnson, P.J., and Yim, A.P. 2003. Genetic alterations of lung adenocarcinoma in relation to smoking and ethnicity. *Lung Cancer* **41**(1): 91-99.

Takahashi, T., Nau, M.M., Chiba, I., Birrer, M.J., Rosenberg, R.K., Vinocour, M., Levitt, M., Pass, H., Gazdar, A.F., and Minna, J.D. 1989. p53: a frequent target for genetic abnormalities in lung cancer. *Science* **246**(4929): 491-494.

Toyooka, S., Tsuda, T., and Gazdar, A.F. 2003. The TP53 gene, tobacco exposure, and lung cancer. *Hum Mutat* **21**(3): 229-239.

Westra, W.H., Slebos, R.J., Offerhaus, G.J., Goodman, S.N., Evers, S.G., Kensler, T.W., Askin, F.B., Rodenhuis, S., and Hruban, R.H. 1993. K-ras oncogene activation in lung adenocarcinomas from former smokers. Evidence that K-ras mutations are an early and irreversible event in the development of adenocarcinoma of the lung. *Cancer* **72**(2): 432-438.

Wistuba, II, Gazdar, A.F., and Minna, J.D. 2001. Molecular genetics of small cell lung carcinoma. *Semin Oncol* **28**(2 Suppl 4): 3-13.

Wong, M.P., Fung, L.F., Wang, E., Chow, W.S., Chiu, S.W., Lam, W.K., Ho, K.K., Ma, E.S., Wan, T.S., and Chung, L.P. 2003. Chromosomal aberrations of primary lung adenocarcinomas in nonsmokers. *Cancer* **97**(5): 1263-1270.

Wong, M.P., Lam, W.K., Wang, E., Chiu, S.W., Lam, C.L., and Chung, L.P. 2002. Primary adenocarcinomas of the lung in nonsmokers show a distinct pattern of allelic imbalance. *Cancer Res* **62**(15): 4464-4468.

You, M.J., Castrillon, D.H., Bastian, B.C., O'Hagan, R.C., Bosenberg, M.W., Parsons, R., Chin, L., and DePinho, R.A. 2002. Genetic analysis of Pten and Ink4a/Arf interactions in the suppression of tumorigenesis in mice. *Proc Natl Acad Sci U S A* **99**(3): 1455-1460.

Zabarovsky, E.R., Lerman, M.I., and Minna, J.D. 2002. Tumor suppressor genes on chromosome 3p involved in the pathogenesis of lung and other cancers. *Oncogene* **21**(45): 6915-6935.