AFFYMETRIX®

The Way Ahead.™

High Resolution Chromosome Copy Number Analysis Using GeneChip® Mapping Arrays

Jeff Smith, Ph.D.
MidAtlantic Genotyping Specialist
Affymetrix

AFFYMETRIX

The Way Ahead.™

Genomic Copy Number Analysis

- Cytogenetics - De novo and inherited germ line changes
- Cancer Research - Somatic amplifications and deletions
- Association studies – inherited germ line changes

AFFYMETRIX®

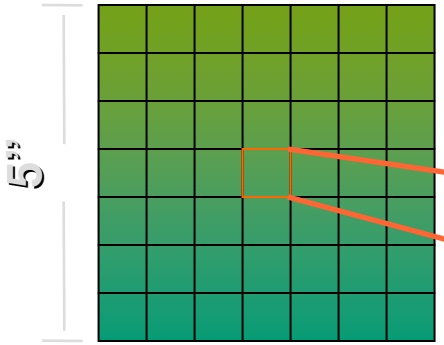
The Way Ahead.™

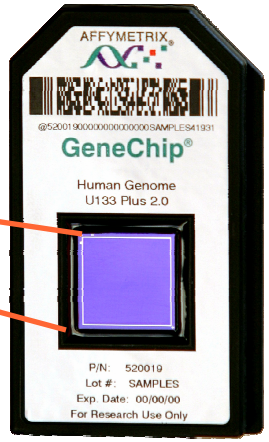
GeneChip® Technology Platform

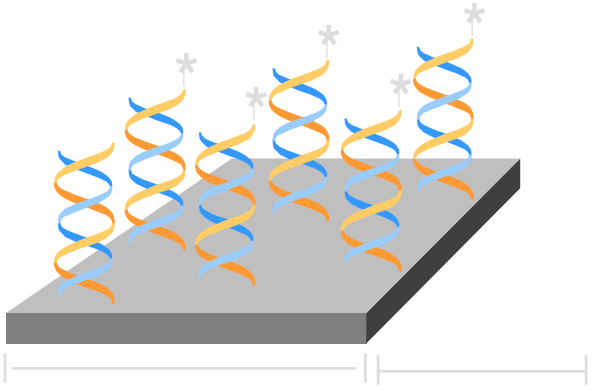
AFFYMETRIX®

The Way Ahead.™

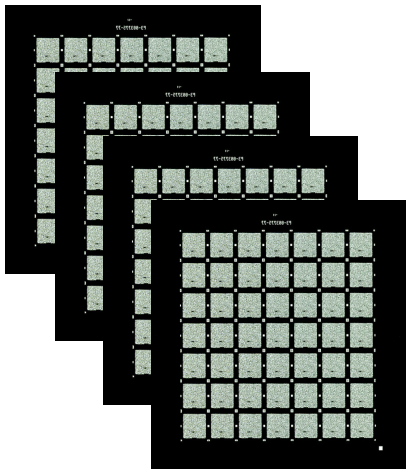
7G GeneChip® Technology: 5µm spacing

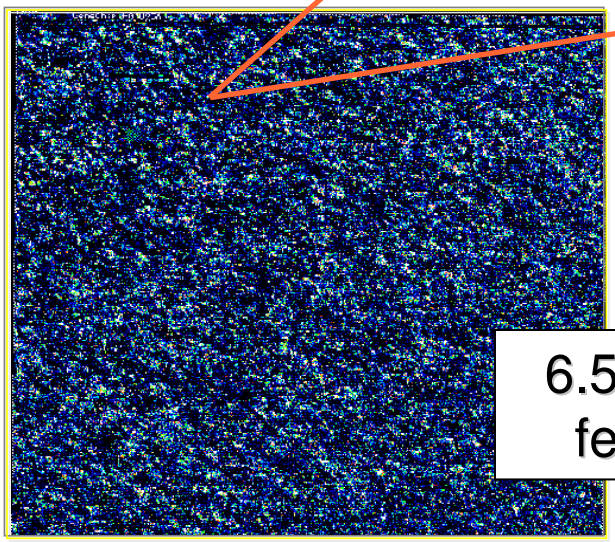






49 Chips per Wafer





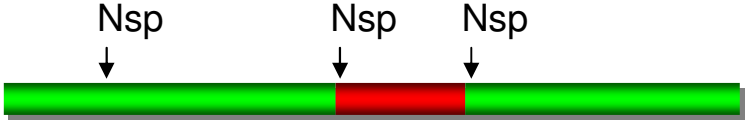
Millions of identical oligos per feature

6.5 Million different features per chip

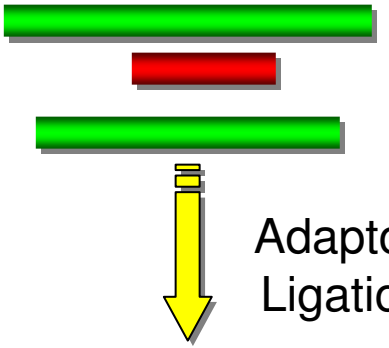
GeneChip® Mapping Assay Overview

AFFYMETRIX
The Way Ahead.™

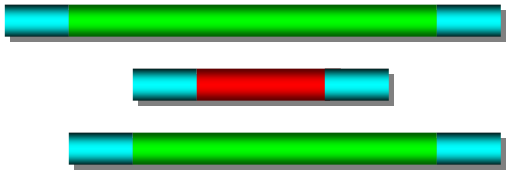
250 ng Genomic DNA



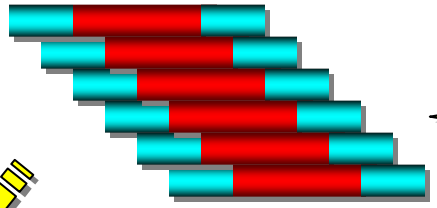
RE Digestion



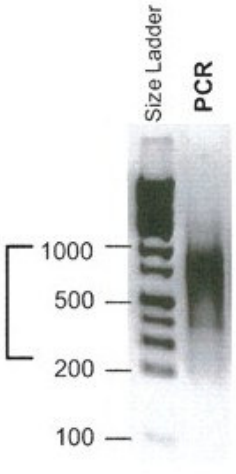
Adaptor Ligation



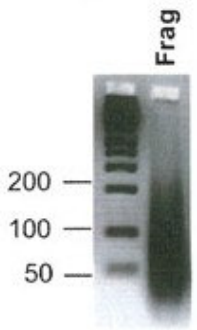
PCR: One Primer Amplification

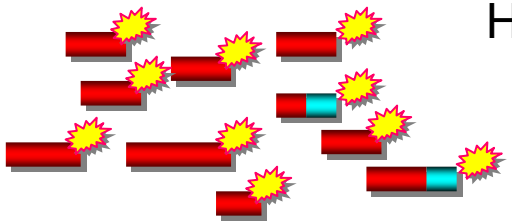


Complexity Reduction

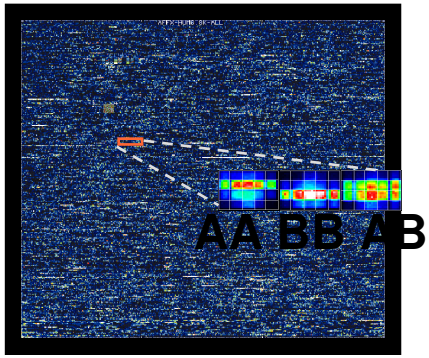


Fragmentation and Labeling



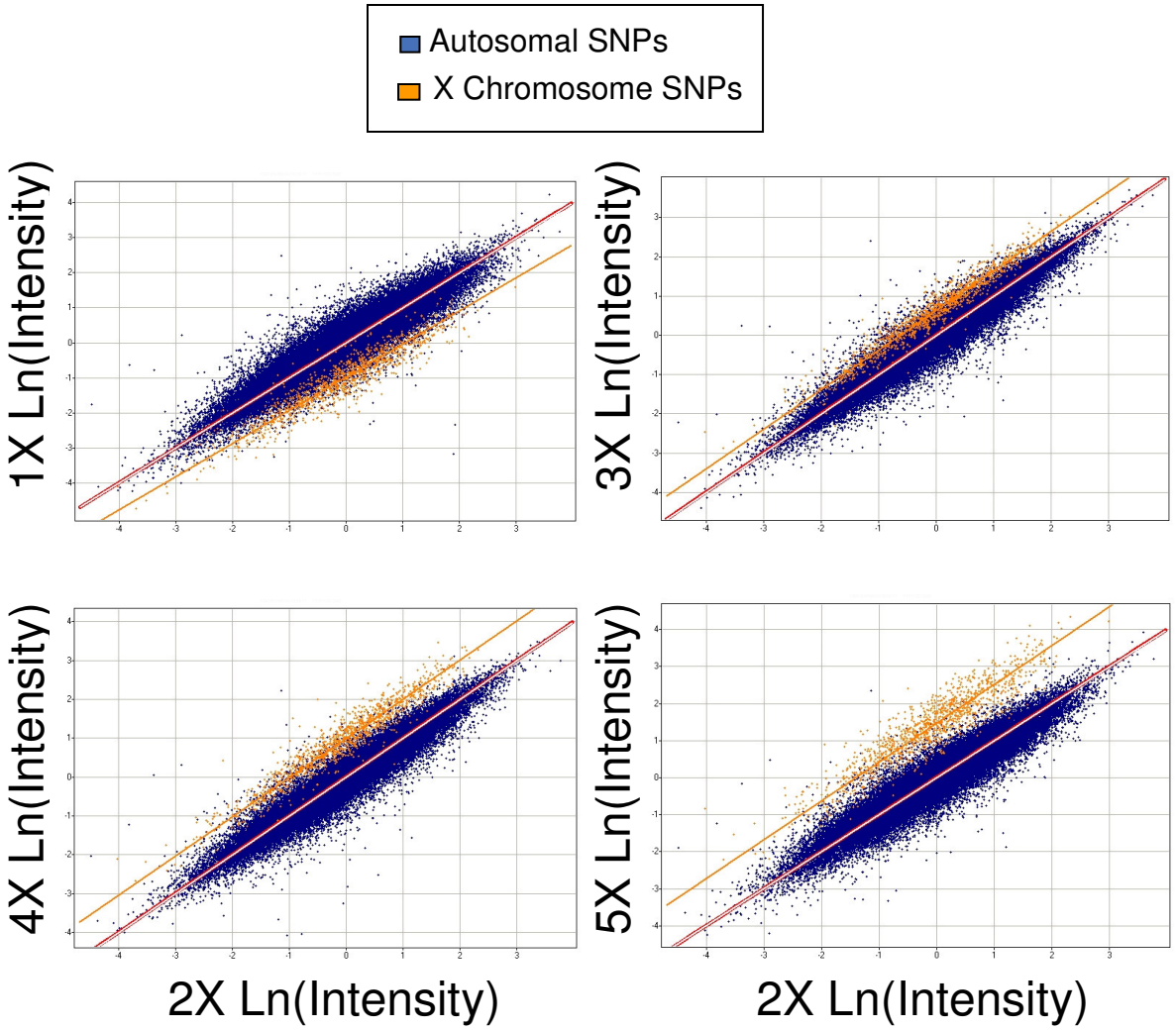


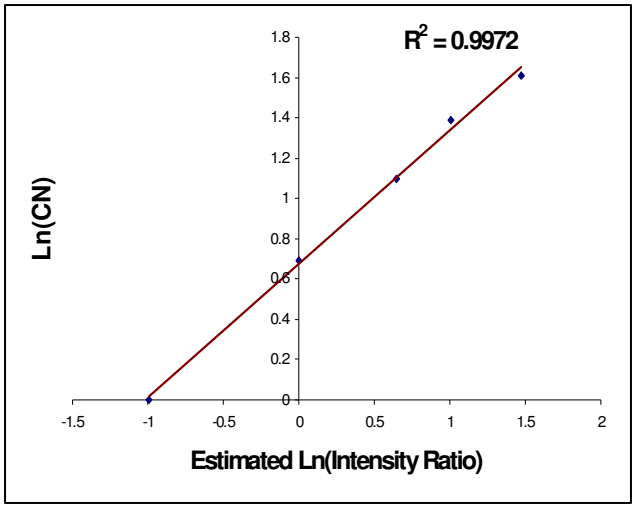
Hyb & Wash



AFFYMETRIX
The Way Ahead.™

SNP Intensity Correlates With Chromosomal Copy Number





(100K data)

AFFYMETRIX

The Way Ahead.™

An Integrated View of Copy Number and Allelic Alterations in the Cancer Genome Using Single Nucleotide Polymorphism Arrays

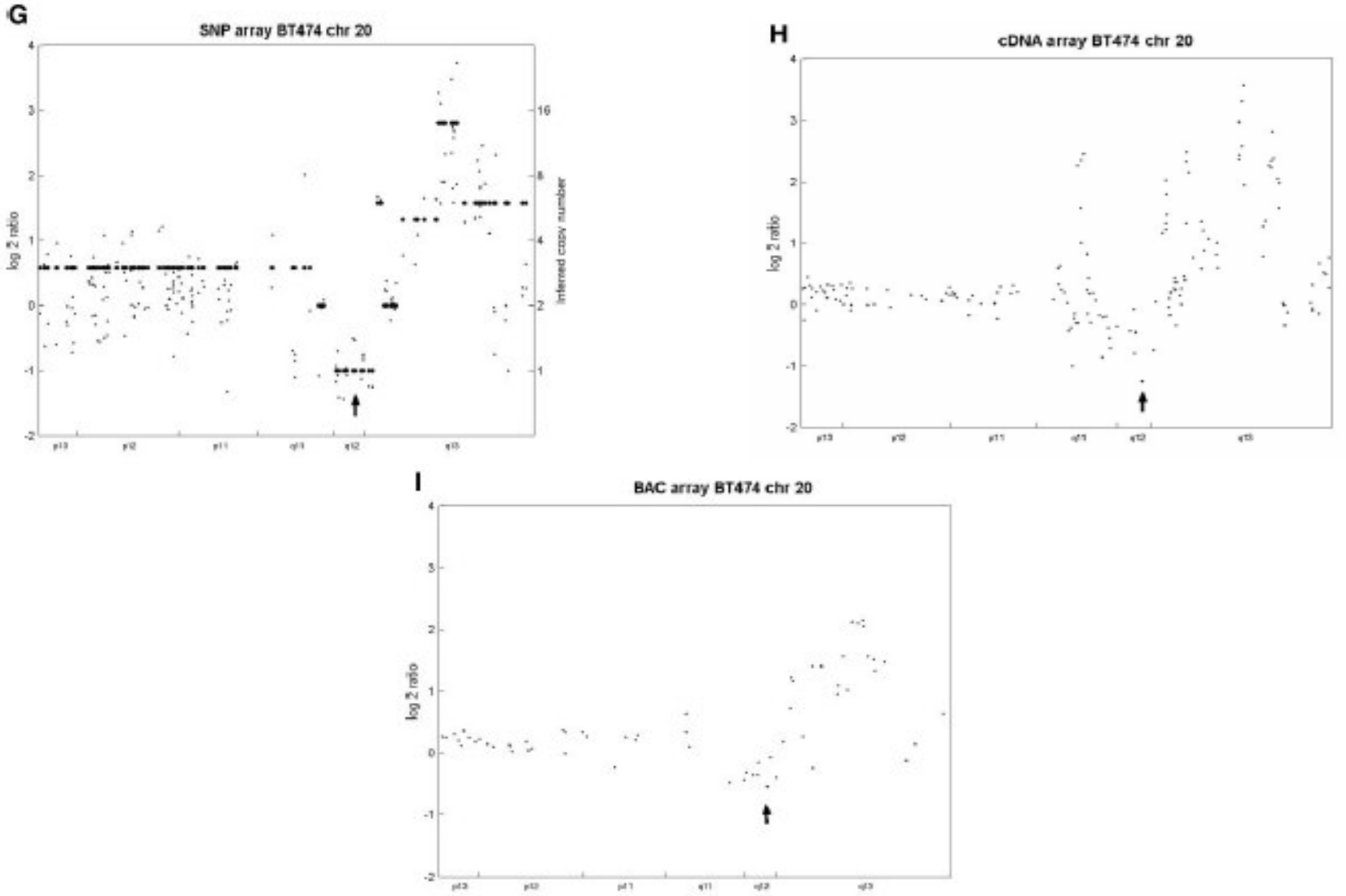
Xiaojun Zhao,[1,4] Cheng Li,[2,5] J. Guillermo Paez,[1,3] Koei Chin,[7] Pasi A. Jänne,[1,3] Tzu-Hsiu Chen,[1] Luc Girard,[8,9] John Minna,[8,9] David Christiani,[6] Chris Leo,[1] Joe W. Gray,[7] William R. Sellers,[1,3] and Matthew Meyerson[1,4]

[1]Departments of Medical Oncology and [2]Biostatistical Sciences, Dana-Farber Cancer Institute, Boston, Massachusetts; [3]Departments of Medicine and [4]Pathology, Harvard Medical School, Boston, Massachusetts; [5]Department of Biostatistics and [6]Environmental Health, Harvard School of Public Health, Boston, Massachusetts; [7]Department of Laboratory Medicine, University of California, San Francisco, California; [8]Hamon Center for Therapeutic Oncology Research, and [9]Departments of Internal Medicine and Pharmacology, University of Texas Southwestern Medical Center, Dallas, Texas

Zhao, X. et al. Cancer Res. 64: 3060 (2004)

AFFYMETRIX
The Way Ahead.™

Comparison of 10K, cDNA, and BAC Arrays – Breast Carcinoma, Chr. 20



Zhao, X. et al. *Cancer Res* 64, 3060-71 (2004).

Copy Number Product Evolution

2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008

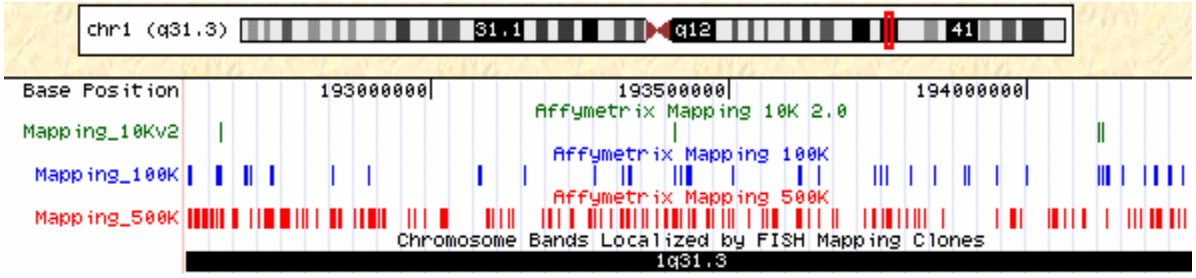
	10K 2.0	100K	500K
Median intermarker distance:	105 kb	8 kb	2.5 kb
Mean intermarker distance:	210 kb	22.5 kb	5.8 kb

HuSNP
LOH Analysis

10K
LOH/ CN Analysis
CNAT1.0 / Affy Tool

100K
LOH/ CN Analysis
CNAT 2.0/3.0

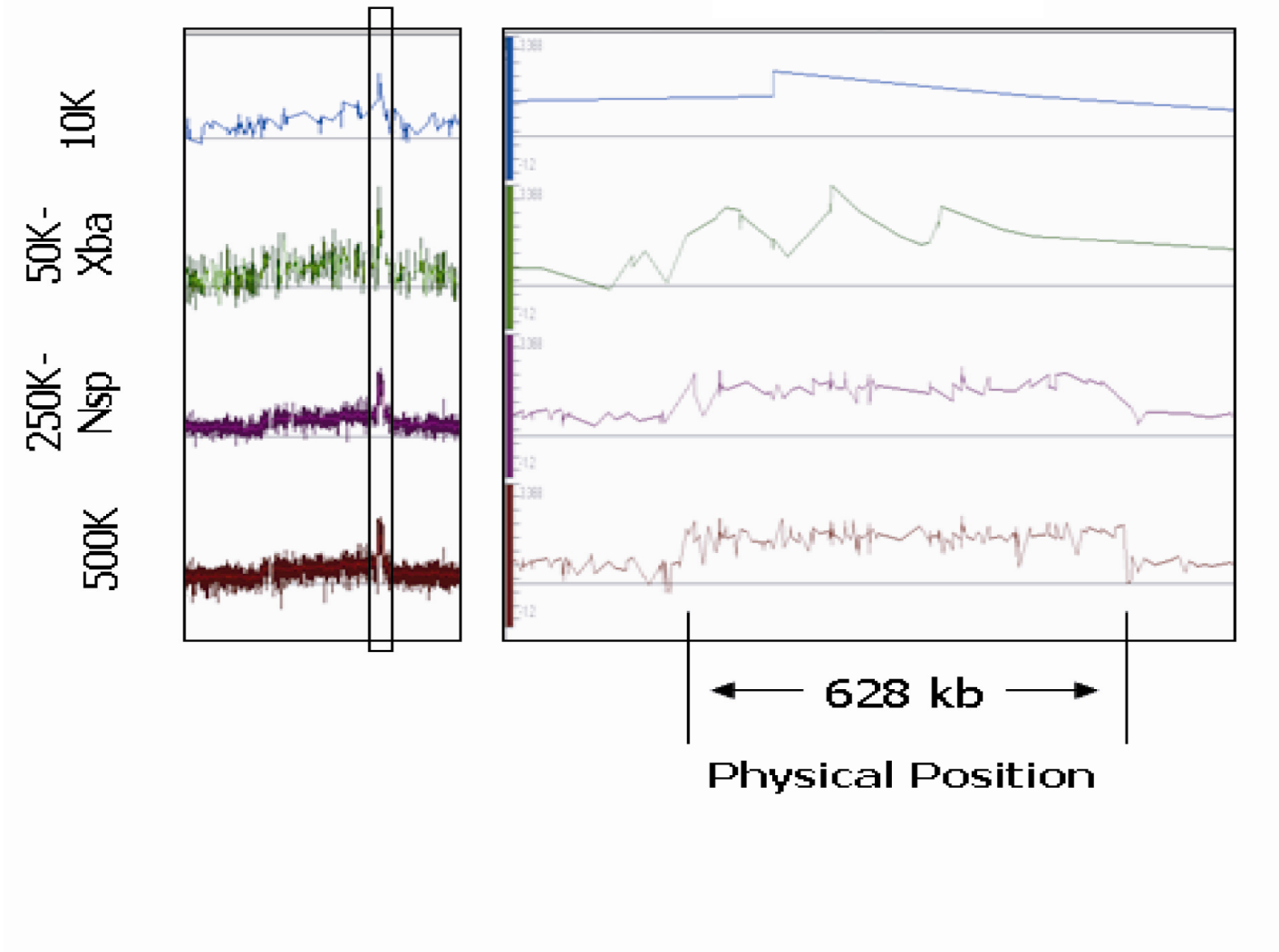
500K
LOH/ CN Analysis
CNAT 4.0



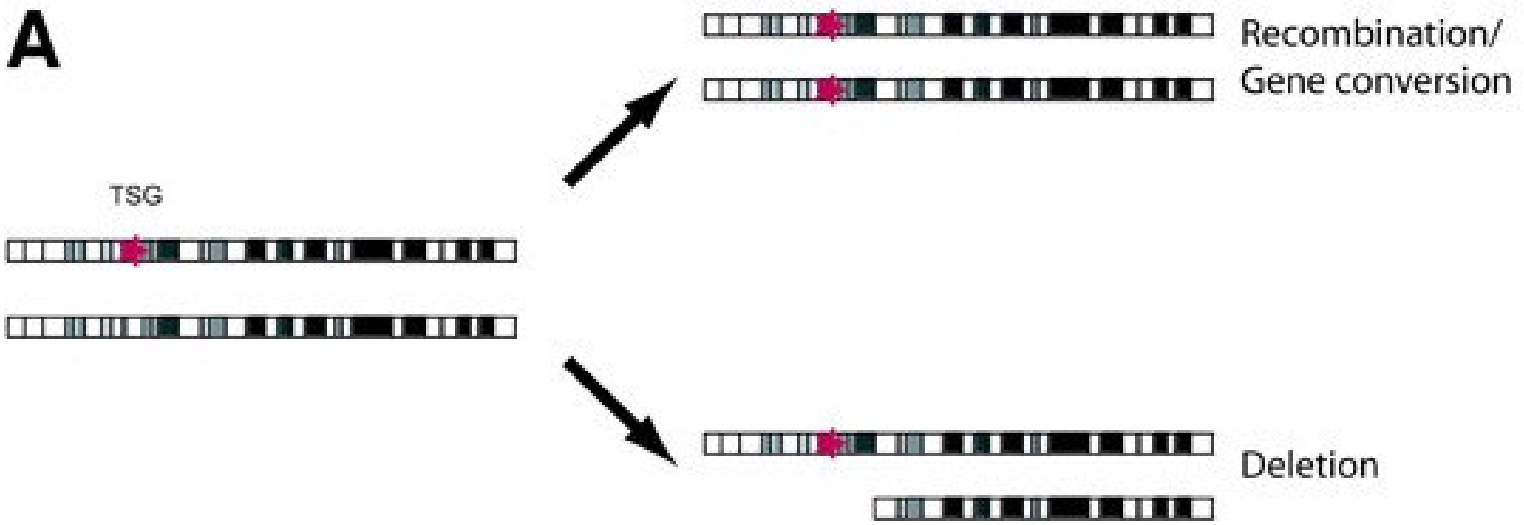
CN Specific Content
CNAT 5.0

AFFYMETRIX
The Way Ahead.™

500K Mapping Arrays Enable Fine-Mapping of Gains and Losses



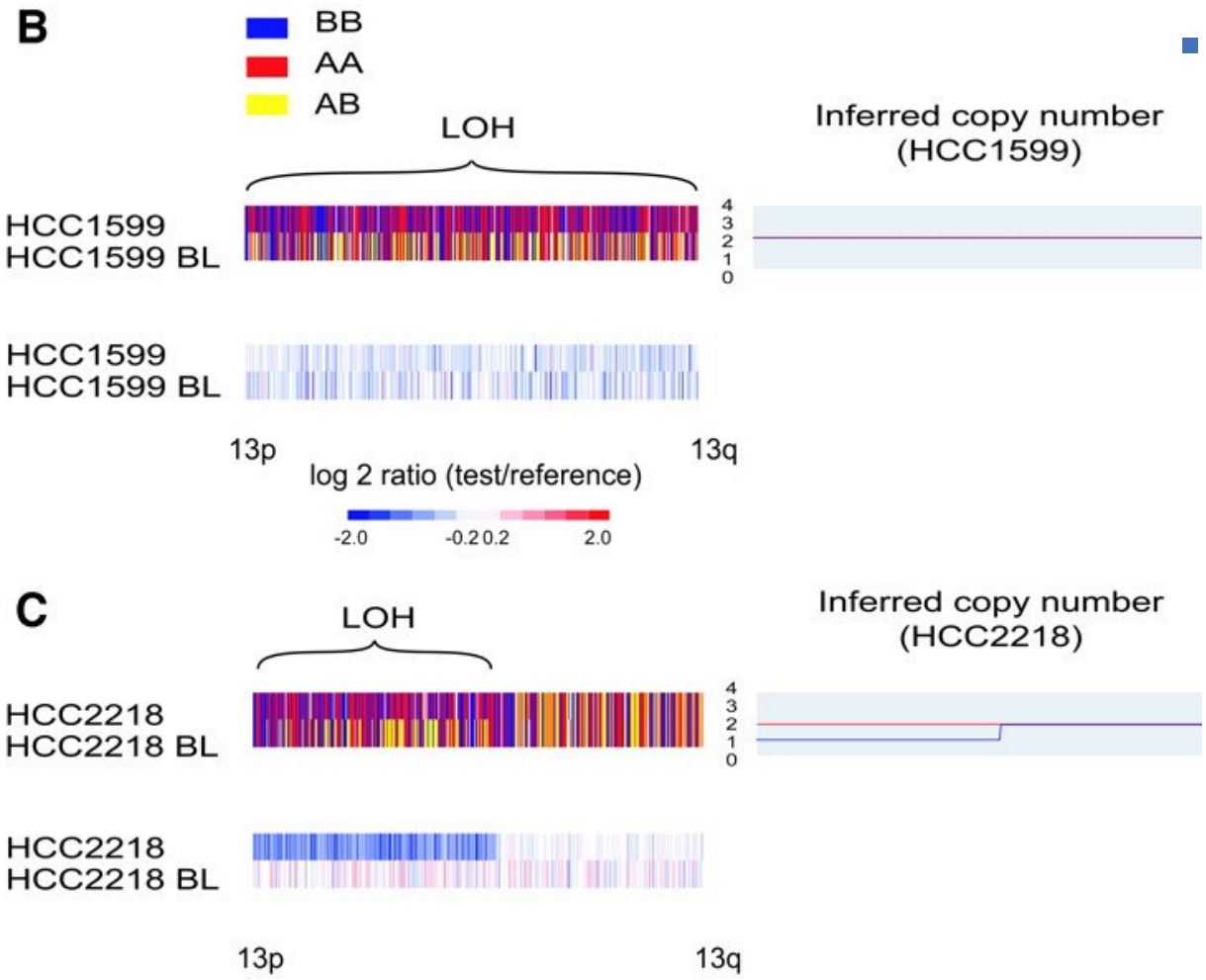
GeneChip® Mapping Arrays Distinguish Between Different LOH Mechanisms



- Different mechanisms can cause LOH (Panel A)
  - Point mutation
  - Hemizygous deletion
  - Mitotic disjunction
  - Mitotic recombination
  - Gene conversion

Zhao, X. et al. *Cancer Res* 64, 3060-71 (2004).

GeneChip® Mapping Arrays Distinguish Between Different LOH Mechanisms



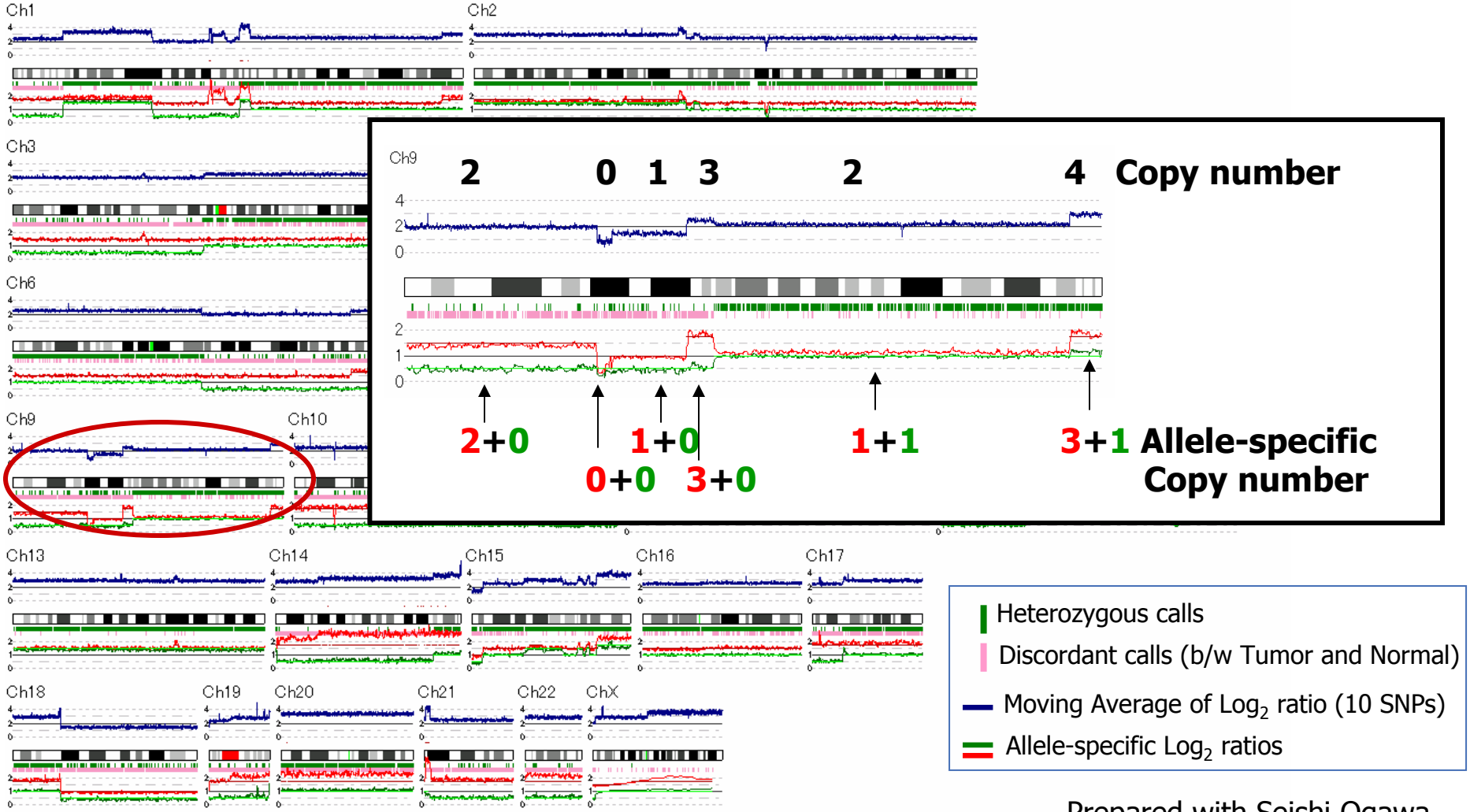
- SNPs can distinguish different genetic mechanisms that lead to LOH (Panel B and C)
  - In HCC1599, chromosome 13 undergoes LOH, as determined by genotype analysis, but no change in copy number (Panel B)
  - In contrast, HCC2218, chromosome 13 undergoes LOH and copy number change

Zhao, X. et al. *Cancer Res* 64, 3060-71 (2004).

Lung Cancer Sample on 500K Allele Specific Copy number Analysis

AFFYMETRIX
The Way Ahead.™

Single sample (CCL-256D) compared to matched normal (CCL-256.1D)



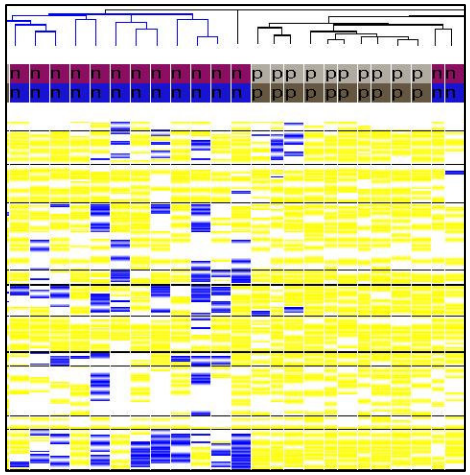
- Heterozygous calls
- Discordant calls (b/w Tumor and Normal)
- Moving Average of Log_2 ratio (10 SNPs)
- Allele-specific Log_2 ratios

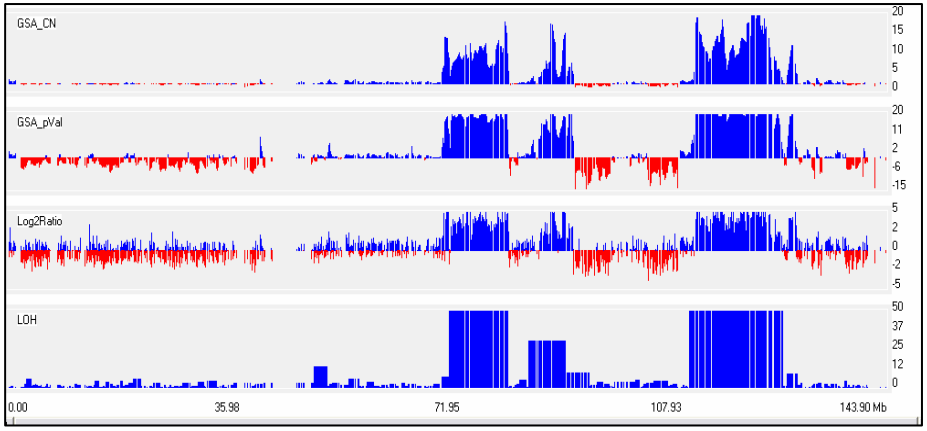
Prepared with Seishi Ogawa
CNAG: www.genome.umin.jp

Affymetrix GeneChip® Mapping Arrays for Genotype, LOH, and Copy Number

- Mapping arrays can be used for combined genotype, loss of heterozygosity, and copy number analysis

SNP ID	Chromosome	Physical Position	dbSNP RS ID	Ref103_2 Call	Ref103_2 Confidence
SNP_A-4286093	16	34679872	rs8054993	BB	0.000488
SNP_A-2131822	10	70635063	rs12256724	BB	0.046875
SNP_A-4265885	6	117421903	rs1413751	AA	0.000488
SNP_A-2057735	16	46397850	rs16945629	BB	0.007813
SNP_A-4274636	10	70632906	rs7905984	BB	0.007813
SNP_A-4266929	8	47677652	rs1474386	AB	0.007813
SNP_A-4294701	14	66038848	rs7146723	AA	0.000977
SNP_A-1881300	5	70963442	rs277979	AB	0.007813





AFFYMETRIX
The Way Ahead.™

Limitations of Current Technologies

- Do not provide genotype information
  - LOH regions resulting from no copy change are not detected
- Limits of resolution, throughput, and manufacturing
  - CGH - BAC arrays are manufactured on-site
    - Potential for spot variation within and between sites
    - Different clones at different sites
    - Cannot be fabricated to industrial standards

Technology	Amount of DNA Required	Assay Time	Resolution	Probe Sequence Dependent
Karyotyping	++++	slow	5-10 Mb	No
FISH	+++	slow	1-5 Mb	Yes
CGH	+++	medium	1 Mb	No

- 10K and 500K can be used with Formalin-Fixed Paraffin-Embedded samples for CCN and LOH
- Imprinting patterns can be identified using Mapping Arrays (Apr 1, 2006)

AFFYMETRIX®

The Way Ahead.™

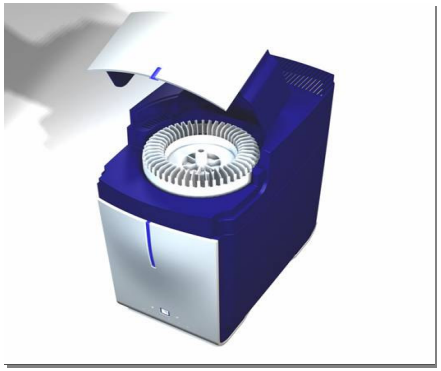
Affymetrix Copy Number Analysis Solutions

GeneChip® Mapping System For Copy Number Analysis

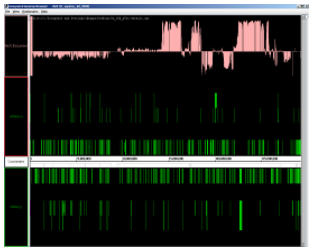
AFFYMETRIX
The Way Ahead.™

GeneChip Mapping Assay Kit w/ Whole Genome Sampling Analysis (WGSA) Assay

GeneChip Mapping 10K, 50K 100K, 500K Arrays

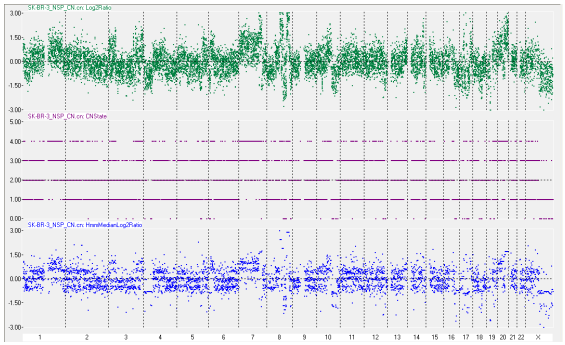


GeneChip Scanner 3000



Integrated Genome Browser

3rd party copy number software tools



Copy Number Analysis Tool 4.0 (CNAT)

- Data Collection
- Data Management
- Client-Server Option

GeneChip Operating Software (GCOS)

AFFYMETRIX
The Way Ahead.™

Data Analysis Solutions

- Affymetrix Copy Number Analysis Tool 4.0
- Academic Tools
  - CNAG
  - GenePattern
  - dChip SNP
- Commercial Solutions- GeneChip Compatible (www.affymetrix.com/products/software/compatible/copynumber.affx)
  - Partek Genomics Suite
  - Stratagene ArrayAssist Copy Number Software
  - Exemplar Copy Number - Sapio Sciences

AFFYMETRIX
The Way Ahead.™

CNAT4.0 Improvements

- Algorithm Improvements –Primary Focus
  - Enable CN and LOH analysis for **500K arrays**
  - Implement **algorithm improvements** to **improve data quality** and **reduce noise** seen with CNATv3
    - Data normalization
      - Probe level- Quantile normalization (e.g.BRLMM)
      - PCR Fragment Size (e.g.CNAG)
      - GC Content (e.g.CNAG).
    - Implement Hidden Markov Model – Defines discrete CN and LOH states
    - Enable **tumor**/ **normal comparison** as well as comparison to a standard or user defined reference set
  - Enable **allele specific analysis** in tumor/ normal analysis
- Software – Incremental Improvements
  - CNAT v3 infrastructure
    - Graphical display of data: Copy Number, Hidden Markov Model States, and Loss of Heterozygosity (LOH)
  - Whole Genome and Chromosome specific views
  - Multi sample view to facilitate identification of trends in copy number or LOH data
  - Dynamic filtering to enable thresholding of data
  - Direct Link to Affymetrix' Integrated Genome Browser (IGB) and easy export to UCSC Browser

AFFYMETRIX
The Way Ahead.™

CNAT 4.0 Key Features/ Benefits

Feature	Benefit
Support for all GeneChip Mapping Products (10K, 50K, 100K, 250K, and 500K arrays)	▪ Higher resolution provides better definition of genetic breakpoints
Allele-specific copy number estimation on matched tumor/normal samples	▪Enables detection of copy neutral events
Copy number and LOH determination using matched normal samples or by using a set of un-matched references	▪Enables comparison of matched tumor/ normal pairs as well as unmatched samples
Whole Genome and Chromosome specific views as well as Multi sample view	▪ to facilitate identification of trends in copy number or LOH data
Dynamic filtering of data	▪Enable the customer to set thresholds and visualize only data meeting that threshold
Export of data in wiggle format for uploading into UCSC browser	▪ Facilitates identification of candidate genes by linking data to genic annotation

AFFYMETRIX
The Way Ahead.™

Overview of Partek

- Partek Genomics Suite is Affymetrix GeneChip-compatible™ for gene expression analysis, exon expression analysis, promoter tiling array analysis, chromosomal copy number analysis, and SNP-based association analysis.
- For Copy Number Analysis:
  - LOH and Copy Number Estimation of imported 10K, 100K, and 500K data
  - Statistical methods for the identification of regions of amplification/deletion
  - Creates SNP or Gene Lists for regions of interest
  - Offers the ability to create diagnostic/prognostic prediction methods
  - Links to NetAffx, IGB, UCSC Genome Browser, GenBank, GEO, and others

AFFYMETRIX
The Way Ahead.™

GenePattern

- **GenePattern** is a powerful analysis workflow tool developed to support **multidisciplinary genomic research programs**.
  - It includes a collection of **analytic and visualization tools**, interfaces for the easy construction of analytic pipelines and integration of new modules without additional programming.
- GenePattern will soon provide the following support for the analysis of SNP microarray data:
  - Scaling of the data to normalize intensity levels across microarray chips.
  - Probe-level modeling to determine an intensity value for each SNP based on the intensity levels of the probes in each probe set.
  - Copy number (CN) calculation to determine the copy number of a target SNP. The calculation, which divides the intensity value of the target SNP by the intensity value of the normal SNP, is also called CN normalization or normalization with respect to normals.
  - Smoothing based on the R package GLAD (Gain and Loss Analysis of DNA), which detects the altered regions in the genomic pattern and assigns a status (normal, gained or lost) to each chromosomal region.
  - Additional analyses to support detection and visualization of LOH and CN alterations.
- Gene Pattern is available for **MacOS, Windows**, and **Linux** platforms

AFFYMETRIX
The Way Ahead.™

Applications of Genomic Copy Number Analysis

- Cancer Research - Somatic amplifications and deletions
- Cytogenetics - De novo and inherited germ line changes
- Association studies – inherited germ line changes

AFFYMETRIX
The Way Ahead.™

Implications of Chromosomal Rearrangements in Human Biology

- Oncogene amplification and/ or the deletion of tumor suppressor genes are hallmarks of cancer initiation and progression
- Copy number changes have recently been implicated in variable response to therapeutic agents.
- The frequency and biological significance of chromosomal mutations may be much greater than previously understood.
- New copy number methodologies are needed to provide
  - Higher marker resolution to facilitate detection of smaller aberrations and fine map boundaries.
  - Genotype information to enable detection of copy neutral events.
  - Ability to support data standards for submission.

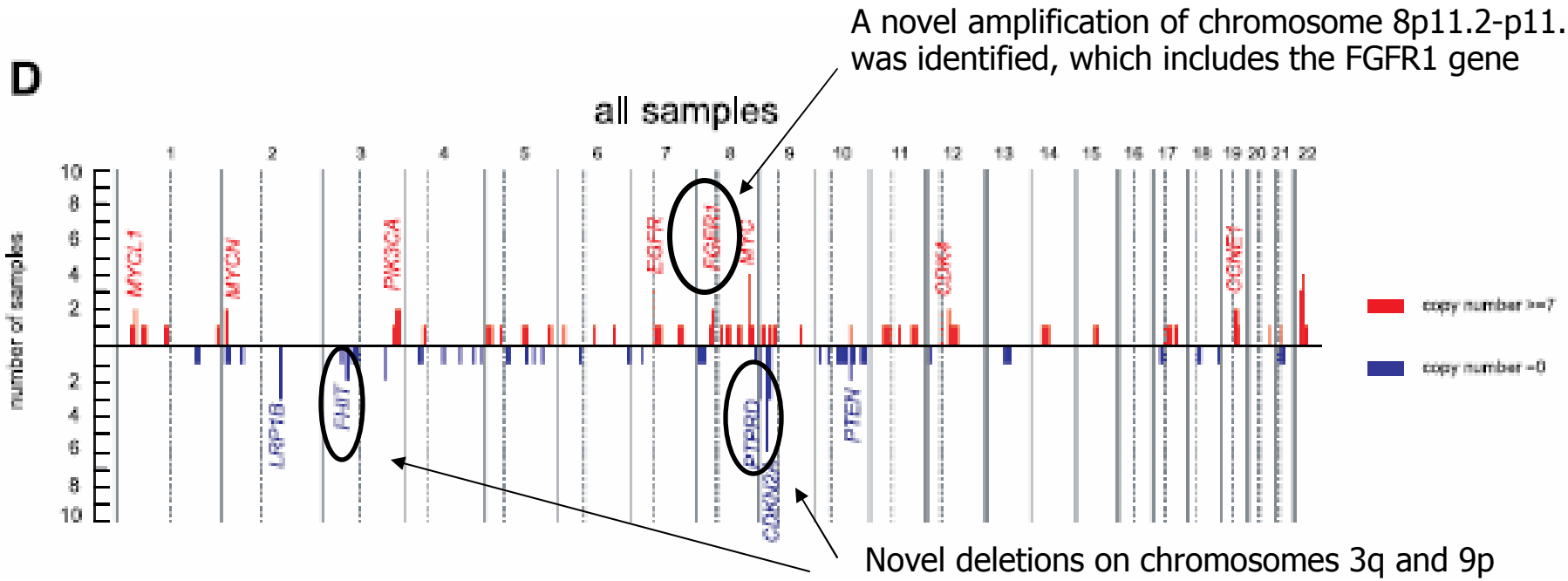
AFFYMETRIX®

The Way Ahead.™

Affymetrix Mapping Arrays for Copy Number Analysis

Cancer Biology

Homozygous Deletions and Chromosome Amplifications in Human Lung Carcinomas Revealed by SNP Array Analysis



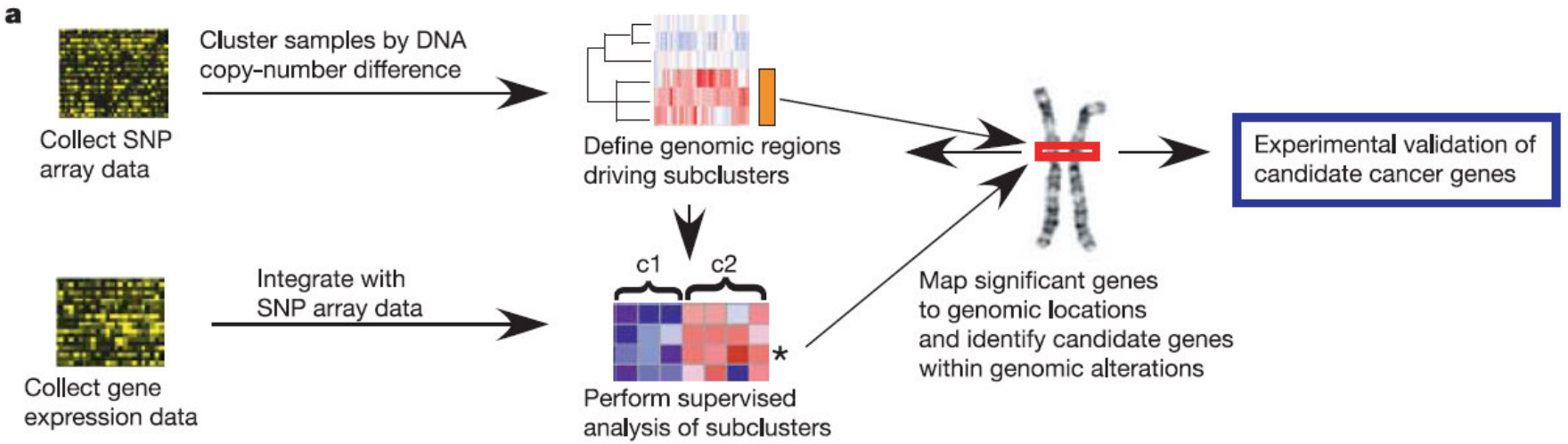
- Red bars – recurrent (≥2) regions of amplification with candidate oncogenes
- Blue bars – recurrent (≥2) homozygous deletions of tumor suppressor gene

Zhao, X. et al Cancer Res. 65: 5561 (2005)

Integration of Expression and Copy Number Data Facilitates Candidate Gene Discovery

Integrative genomic analyses identify *MITF* as a lineage survival oncogene amplified in malignant melanoma

Levi A. Garraway[1,3], Hans R. Widlund[1], Mark A. Rubin[2,3], Gad Getz[5], Aaron J. Berger[6], Sridhar Ramaswamy[5,7], Rameen Beroukhim[1,3], Danny A. Milner[2,3], Scott R. Granter[2], Jinyan Du[1,5], Charles Lee[2,3], Stephan N. Wagner[8], Cheng Li[1,4], Todd R. Golub[1,3,5], David L. Rimm[6], Matthew L. Meyerson[1,2,5], David E. Fisher[1,3] & William R. Sellers[1,2,3,5]



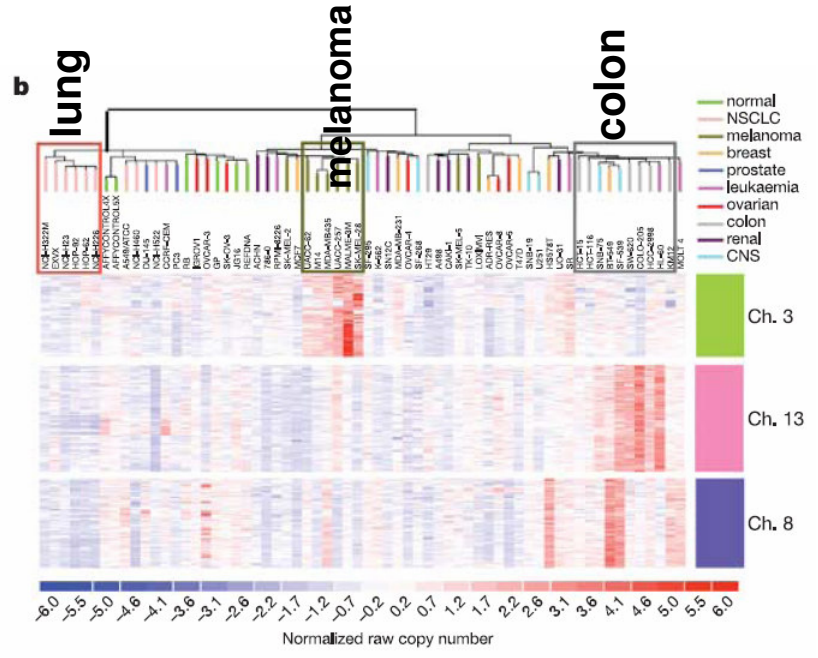
- Performing copy number and expression on the same samples on the same platform provides complementary information on a whole genome scale giving better insight into your biological system

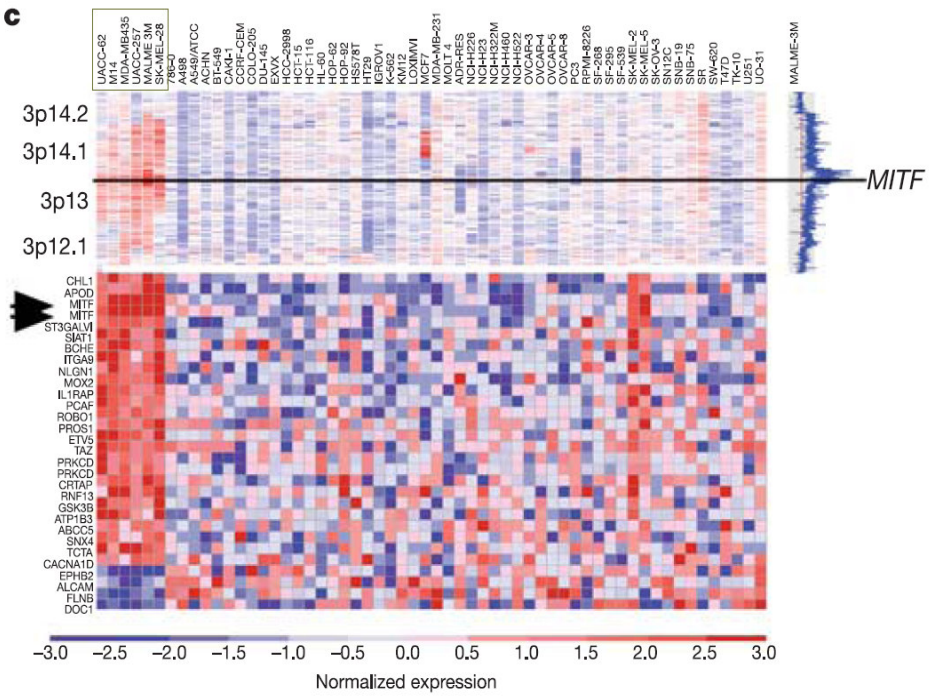
Garraway, L.A. et al Nature 436: 117 (2005)

AFFYMETRIX

The Way Ahead.™

Increased MITF Expression Associated with Amplification in Melanoma Cell Lines using 100K Array





- Hierarchical clustering of raw copy-number data from NCI60 cell lines and normal diploid controls
- Integration of copy-number and gene-expression data

Garraway, L.A. et al Nature 436: 117 (2005)

AFFYMETRIX®

The Way Ahead.™

Affymetrix Mapping Arrays for Copy Number Analysis

Cytogenetics

AFFYMETRIX
The Way Ahead.™

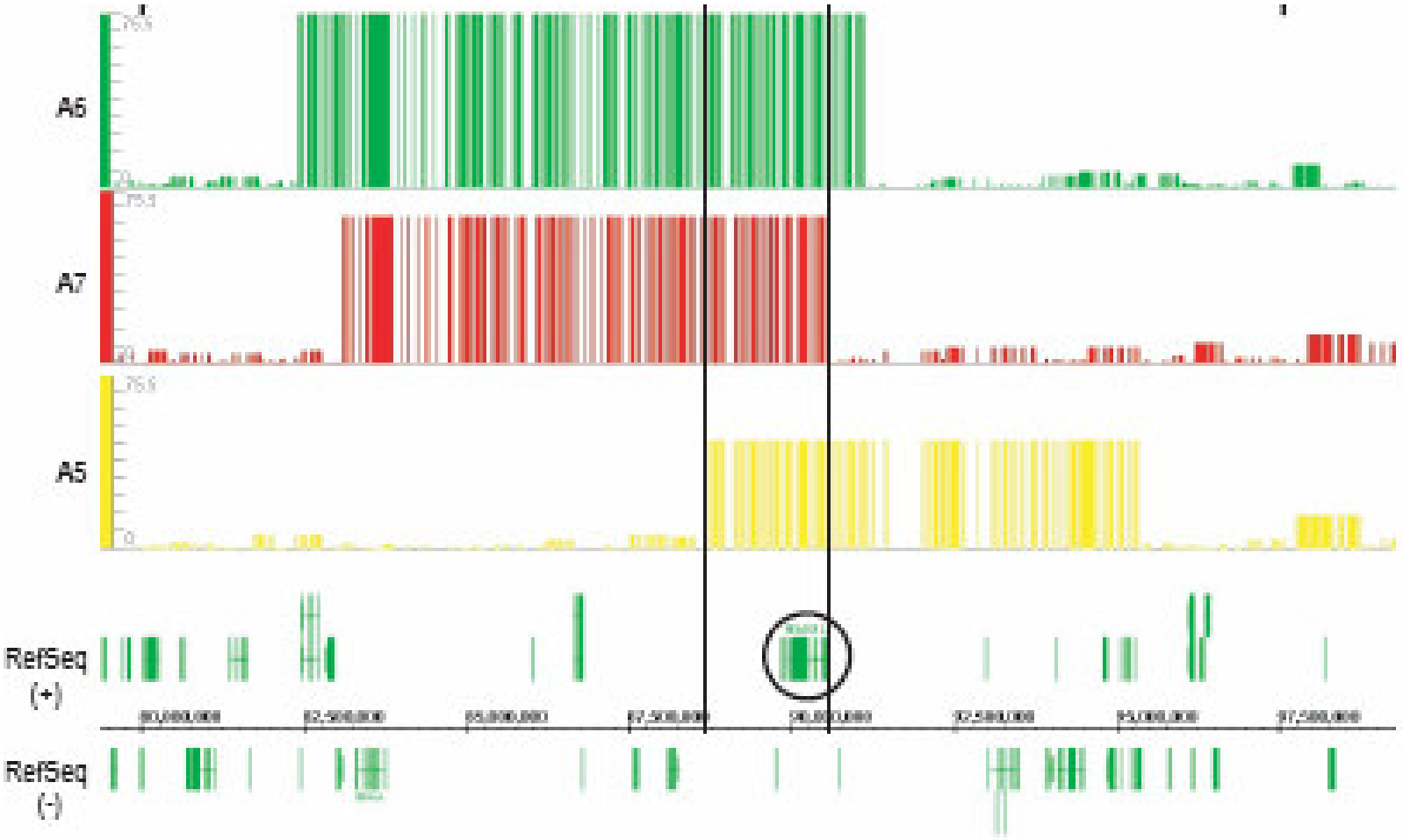
Analysis of Clinical Samples Identifies Gene Associated with Disease

- Set of 3 cases
  - Phenotype – myochronic epilepsy
- Cytogenetics shows small deletion in middle of long arm of chromosome 5
- Three cases are microscopically indistinguishable
- All three patients exhibit seizures but show considerable phenotypic variation

Slater et al., AJHG 2005

AFFYMETRIX
The Way Ahead.™

LOH View of Three Patients on Mapping 100K

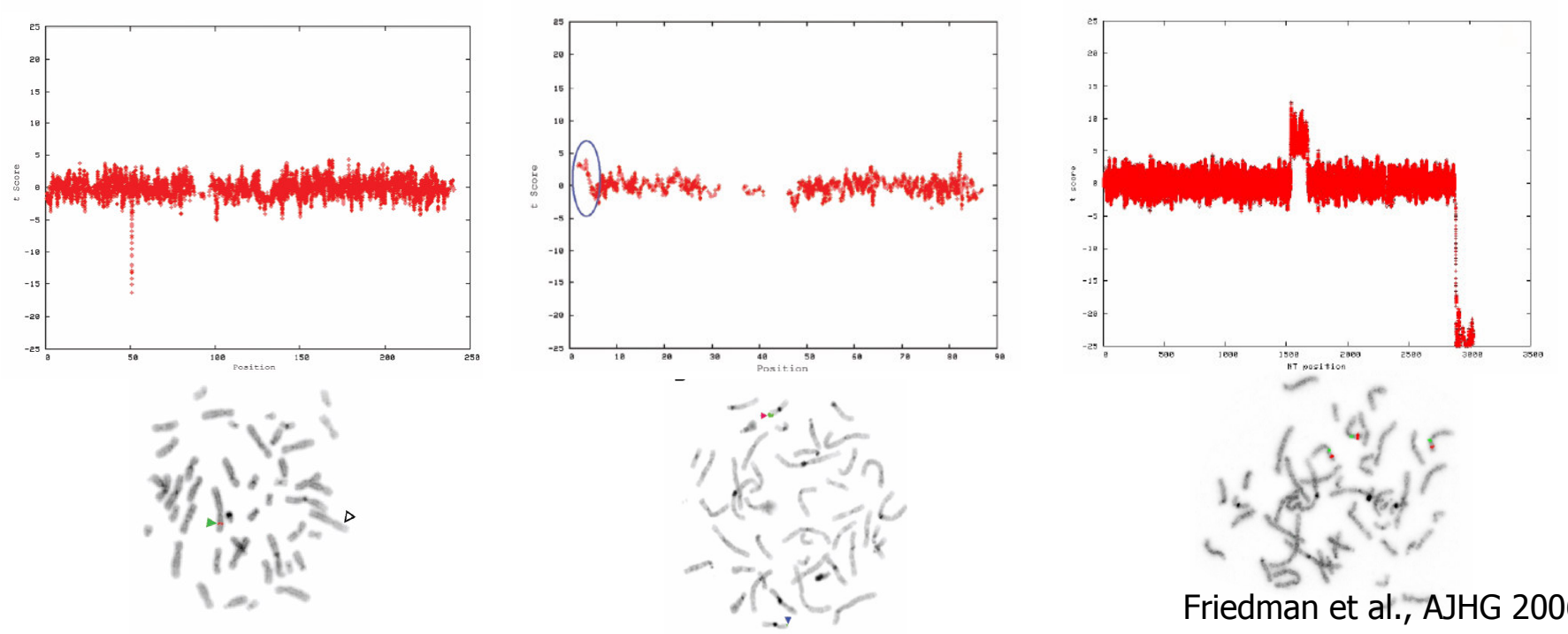


- MASS1 - protein implicated in epileptic disorders

Slater, et al., AJHG 2005

de novo CNVs Identified by the Mapping 100K arrays

- Eleven *de novo* CNVs identified in 100 MR children that appeared normal by traditional cytogenetics
  - 8 submicroscopic deletions; 178 kb to > 11 Mb
  - 2 submicroscopic duplications; 1.1 & 2.9 Mb
  - 1 mosaic trisomy 9
  - None were detected in unaffected siblings suggesting that the *de novo* CNVs likely to pathogenic



Friedman et al., AJHG 2006

AFFYMETRIX®

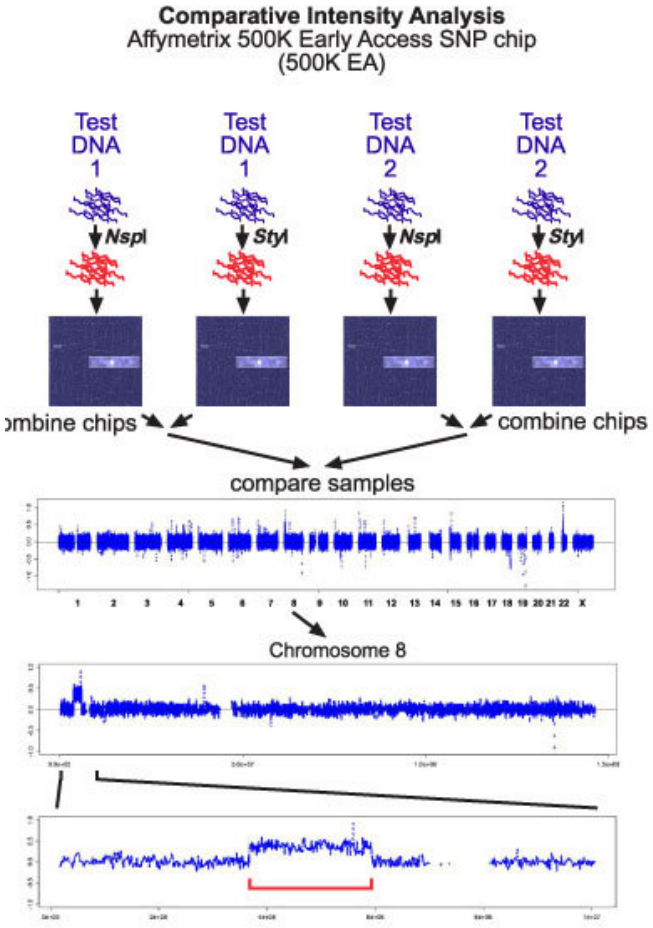
The Way Ahead.™

Affymetrix Mapping Arrays for Copy Number Analysis

Association Studies

Affymetrix 500K Array Used to Generate First Copy Number Variation Map of the Human Genome - November 27, 2007

Global variation in copy number in the human genome. Nature 2006 Nov 23; 444: 444-454.



- **Characterized the frequency of germ line copy number variation (CNV) in the global population (270 HapMap samples) and constructed a first generation CNV map using 500K EA arrays**
- **Key Conclusions**
  - **CNVs are important in human diversity and evolution.** CNV regions encompass more nucleotide content than SNPs. 1447 CNV regions covering 360MB (12% of the genome) were identified in the HapMap populations.
  - **CNVs are important in human biology and disease.** CNV regions contained hundreds of genes (about 2900), functional elements and segmental duplications. Of the 2900 genes within CNV, 285 previously associated with disease.
  - **CNVs are valuable in genetic studies and provide complementary information to SNPs.** Dramatic variation in CNV patterns between HapMap populations and distinct linkage disequilibrium patterns exists for many identified CNVs.
  - ***In combination with SNP information, "CNV assessment should now become standard in the design of all studies of the genetic basis of phenotypic variation, including disease susceptibility."***

AFFYMETRIX
The Way Ahead.™

Whole Genome Copy Number Analysis Using 500K Mapping Arrays

- Increased density of markers provides higher resolution to detect and map boundaries of copy number changes
- Simultaneous collection of genotypes and copy number data allow for:
  - LOH clustering of samples
  - Distinguishing between different mechanism of LOH
  - Detection of copy neutral events, i.e. Uniparental disomy
- Combining of expression and copy number data facilitates identification of novel candidate genes

AFFYMETRIX®

The Way Ahead.™

New Products from Affymetrix for CNV-based Association Studies

AFFYMETRIX
The Way Ahead.™

The New Whole-Genome Human SNP 5.0

- Single Array configuration of the Mapping 500K Array Set
- Developed in collaboration with the Broad Institute of Harvard and the Massachusetts Institute of Technology
- In addition to SNP's, CNV's offer another set of markers that can be used to identify genetic associations with disease and basic human variability
- Consists of:
  - Approximately 500K SNPs from Mapping 500K
  - All SNPs back-compatible
  - 500,000 non-polymorphic tiling probes for the detection of CNV's (100,000 designed in 2,000 known regions of CNV)
- $250/sample
- Available in February 2007
- Copy Number Tool available from the Broad Institute mid-2007

AFFYMETRIX
The Way Ahead.™

The New Whole-Genome Human SNP 6.0

- Single Array configuration of the Mapping 500K Array Set including an additional ~500K HapMap and non-HapMap SNPs
- Developed in collaboration with the Broad Institute of Harvard and the Massachusetts Institute of Technology
- Consists of:
  - Approximately 500K SNPs from Mapping 500K
  - All SNPs back-compatible
  - Approximately 500K additional SNPs to boost genomic coverage for SNP-based association studies
  - Yet to defined number of non-polymorphic tiling probes for the detection of CNV's (100,000 designed in 2,000 known regions of CNV)
- Expected price – Less than $500/sample
- Available in July 2007
- Copy Number Tool available from the Broad Institute mid-2007

AFFYMETRIX
The Way Ahead.™

Whole Genome Copy Number Analysis Using 500K Mapping Arrays

- Increased density of markers provides higher resolution to detect and map boundaries of copy number changes
- Simultaneous collection of genotypes and copy number data allow for:
  - LOH clustering of samples
  - Distinguishing between different mechanism of LOH
  - Detection of copy neutral events, i.e. Uniparental disomy
- Combining of expression and copy number data facilitates identification of novel candidate genes

The Way Ahead.™