AFFYMETRIX®

The Way Ahead.™

Whole-Genome Association using Mapping 500K Array Products

Jeff Smith, Ph.D.
MidAtlantic Genotyping Specialist
Affymetrix

AFFYMETRIX
The Way Ahead.™

Affymetrix Data Analysis Workshop – Whole Genome Association

- Whole Genome Association is in its infancy.
  - Hundreds of thousands of samples ascertained worldwide
  - High throughput genotyping is available at low cost and high accuracy
- Many key questions:
  - Experimental design – SNP selection strategy, power calculations, phenotype selection
  - Data management – storage, sharing, filtering
  - Data analysis- multiple testing, replication strategy, false positives and false negatives
  - Copy number and disease: CNVs, disease associated deletions and amplifications

AFFYMETRIX

The Way Ahead.™

Examples of validated associations using Affymetrix Mapping Arrays

- Complement Factor H for age-related macular degeneration (Hoh et al, *Science*, March 2005)
- Capon for QT intervals (Arking & Chakravarti, ASHG 2005) – Nature Genetics, April 2006
- Insig2 for obesity in Framingham Heart Study (Herbert, ASHG 2005) – Science. April 14
- Myocardial infarction (Stoll, ASHG 2005 – using 500K)
- *KIBRA* association with Memory Performance (Papassotiropoulos, et. al. Science. 20 October 2006)

AFFYMETRIX®

The Way Ahead.™

GeneChip® Technology Platform

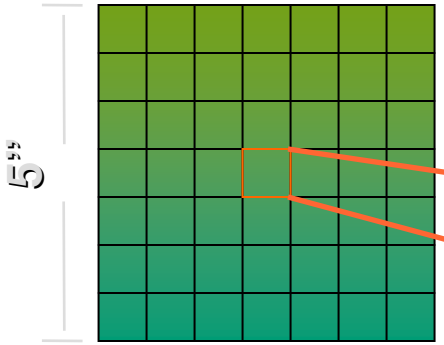
Science is getting more powerful

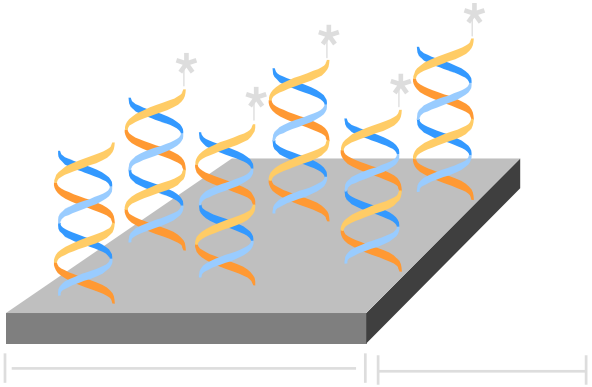
Data sets are getting bigger and better

AFFYMETRIX®

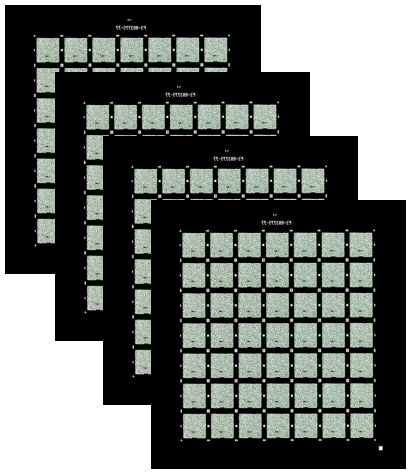
The Way Ahead.™

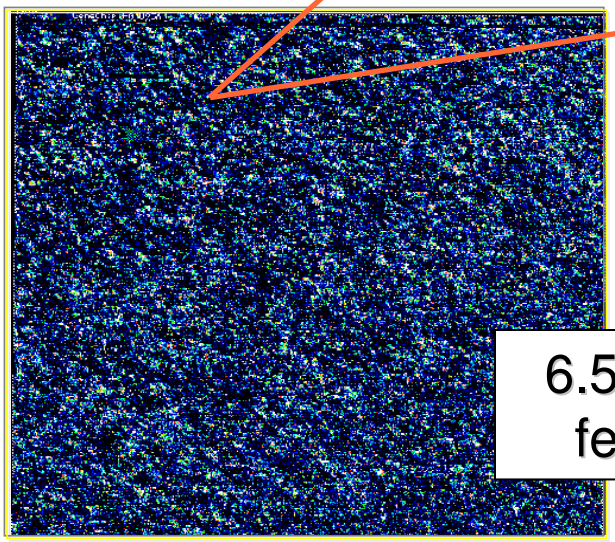
7G GeneChip® Technology: 5µm spacing





49 Chips per Wafer

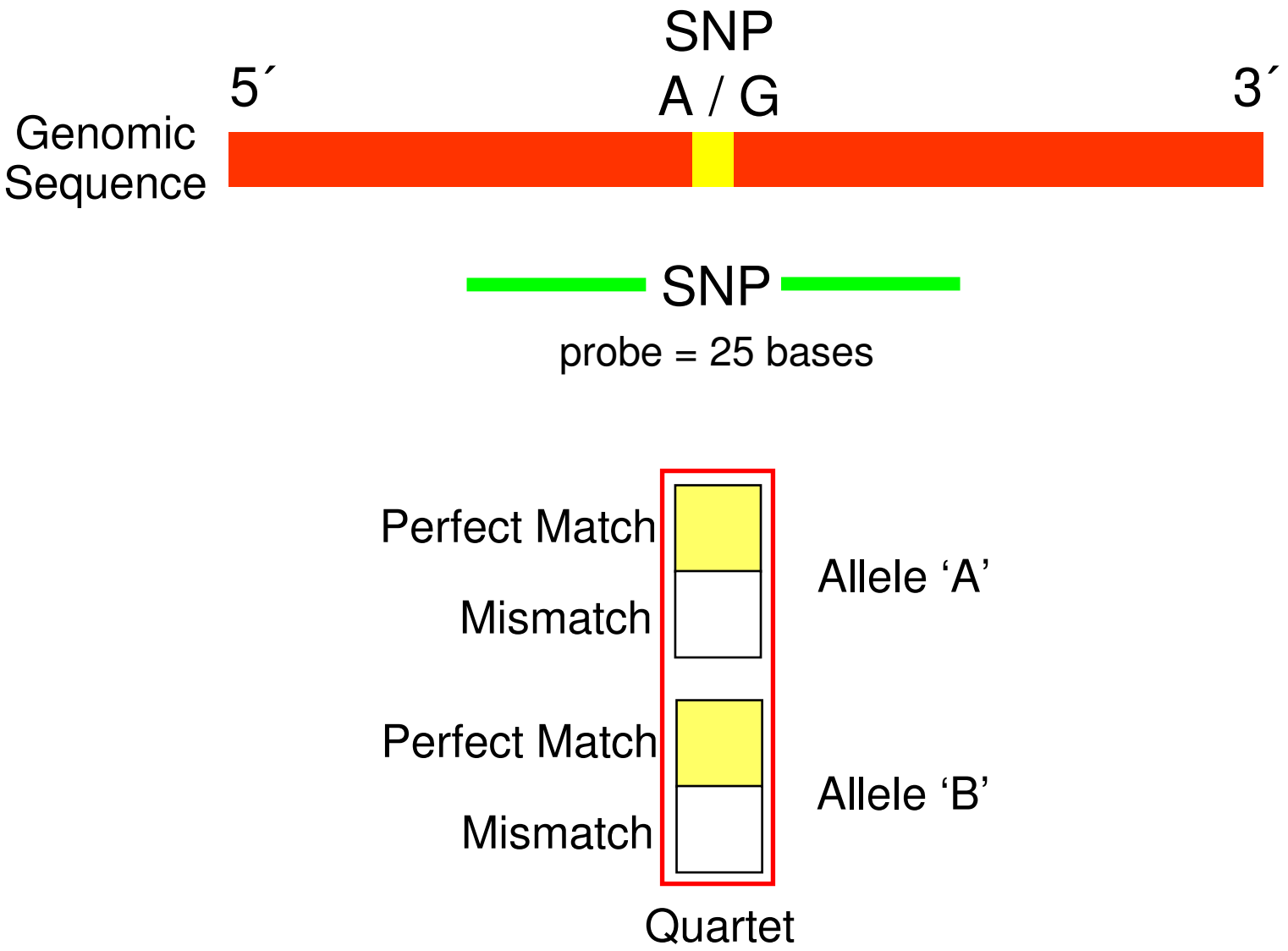




Millions of identical oligos per feature

6.5 Million different features per chip

GeneChip® Mapping Product Design



AFFYMETRIX
The Way Ahead.™

GeneChip Mapping Product - Probe Array Tiling

-2	-1	0	+1	+2	+4
PMA	PMA	PMA	PMA	PMA	PMA
MMA	MMA	MMA	MMA	MMA	MMA
PMB	PMB	PMB	PMB	PMB	PMB
MMB	MMB	MMB	MMB	MMB	MMB

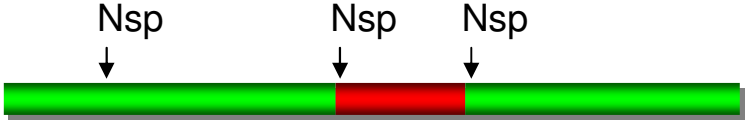
Quartet

- Multiple Quartets are evaluated per SNP

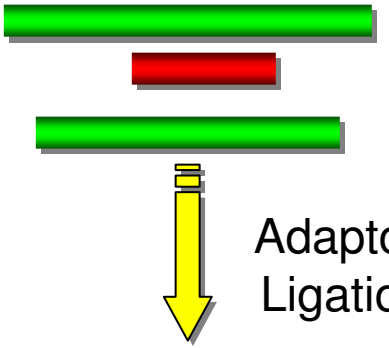
GeneChip® Mapping Assay Overview

AFFYMETRIX
The Way Ahead.™

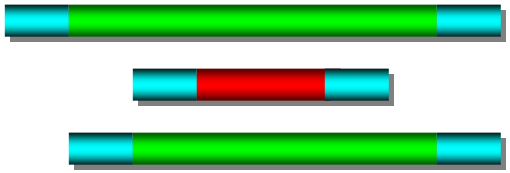
250 ng Genomic DNA



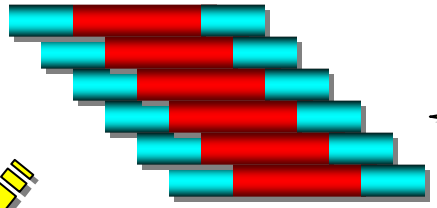
RE Digestion



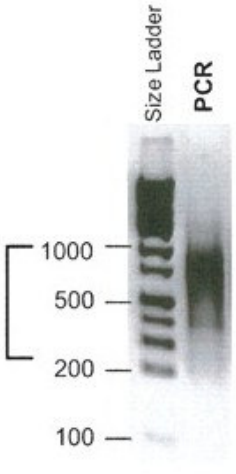
Adaptor Ligation



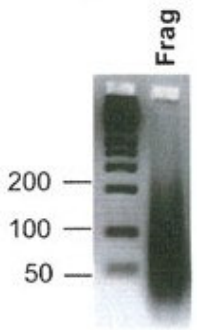
PCR: One Primer Amplification

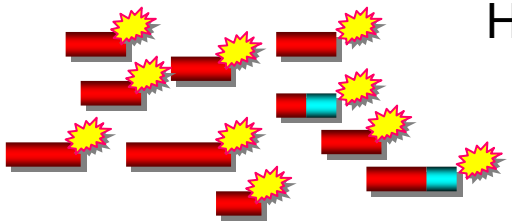


Complexity Reduction

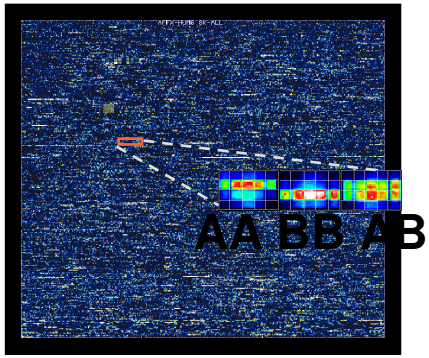


Fragmentation and Labeling





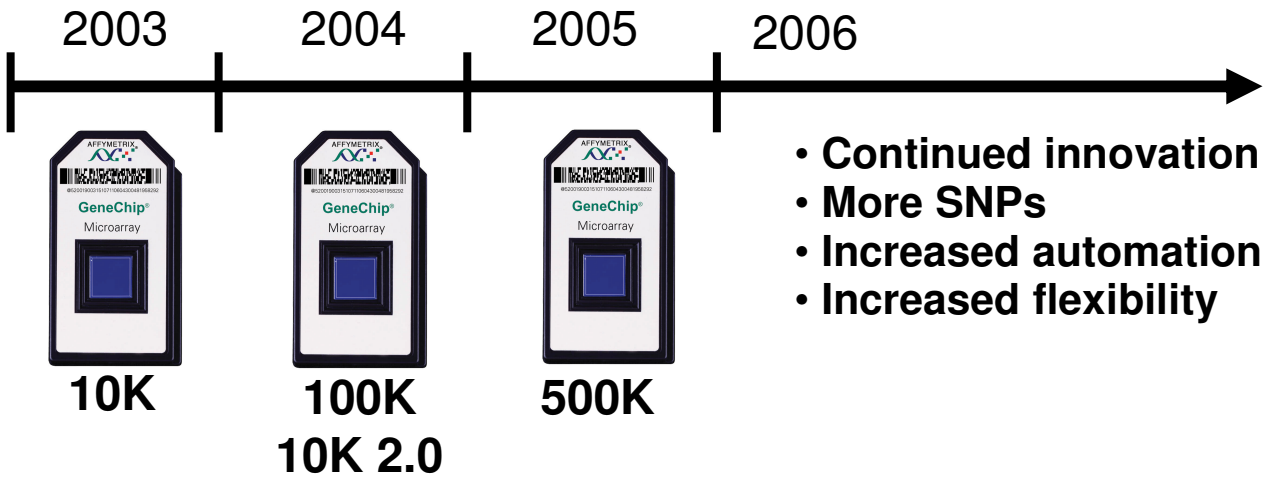
Hyb & Wash



AFFYMETRIX
The Way Ahead.™

Probe and Feature Reduction Drives More Information

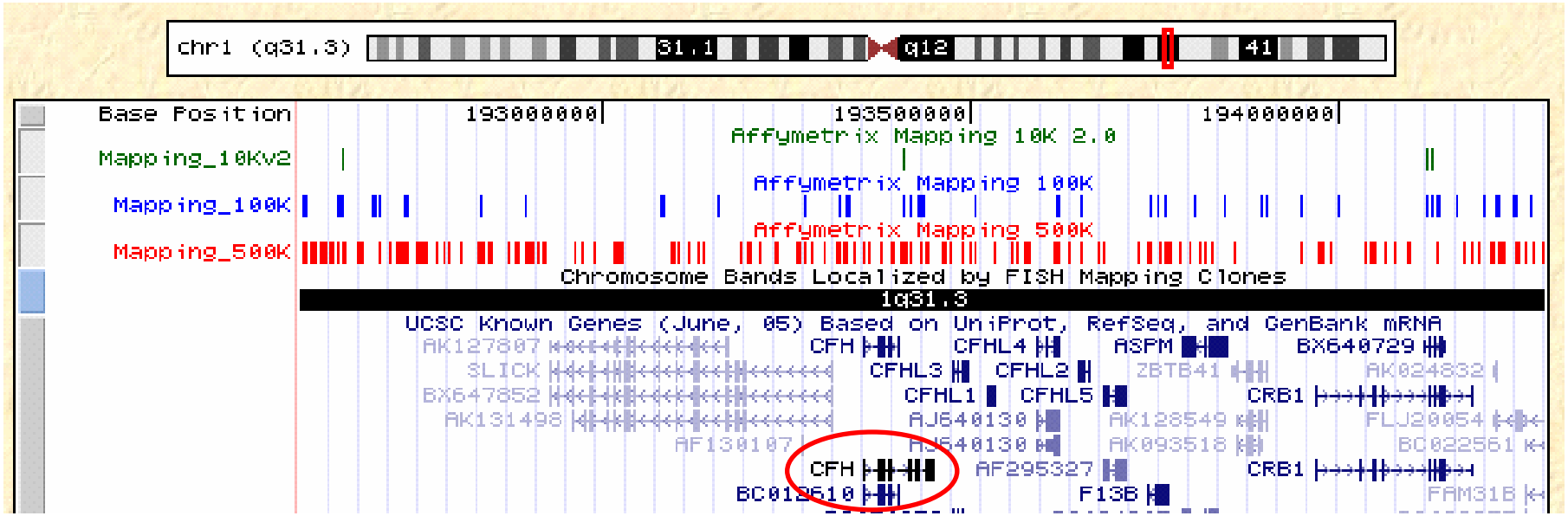
Feature Size	Number of Probes/SNP	Number of SNPs/array
18um	40	**11,500**
8um	40	**60,000**
5um	24	**268,000**
5um	12	**500,000**



AFFYMETRIX

The Way Ahead.™

Mapping Assays: Proven Scalability and Performance



AFFYMETRIX®

The Way Ahead.™

GeneChip® Mapping 500K Array Set

GeneChip® Mapping 500K Array Set

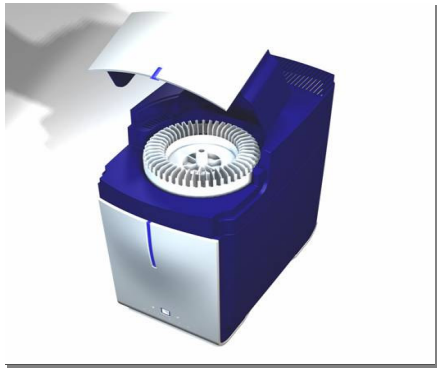
- 4th in the Affymetrix Mapping Array Series
  - Launched Late 2005
  - Followed 10K, 10K v2.0, 100K Set in 2004
- Leading Real-World Capability
  - 2-chips with over 500,000 SNPs
  - Well over 250,000 Mapping 100K and 500K arrays shipped to hundreds of customers to date
- New — Updated assay protocol and standardization
  - New training programs
- New — Updated genotype calling methods

GeneChip® 500K Mapping System

AFFYMETRIX
The Way Ahead.™

GeneChip® Mapping Assay Kit

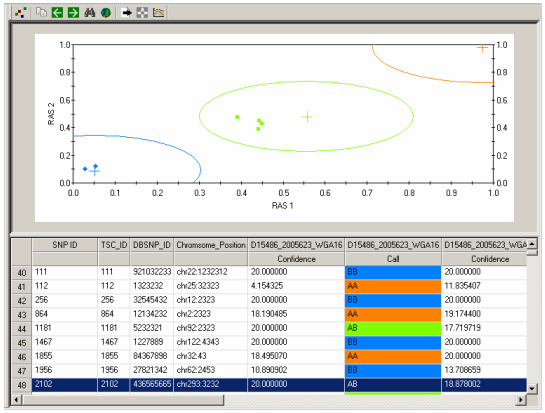
GeneChip® Mapping 500K Set



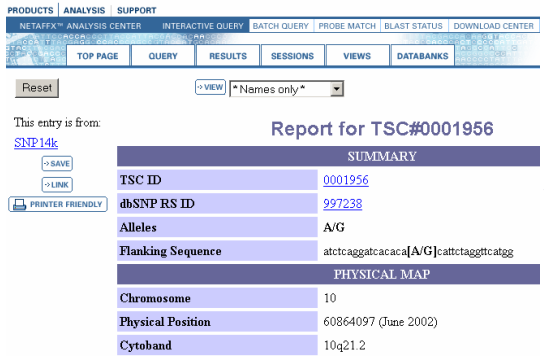
GCS 3000 7G

- Data Collection
- Data Management
- Client-Server Option

GCOS 1.4



GTYPE 4.1
BRLMM Analysis Tool



NetAffx™ SNP Annotation
Downstream Analysis

AFFYMETRIX®
The Way Ahead.™

BRLMM Algorithm (Bayesian Robust Linear Model with Mahalanobis Distance)

- BRLMM delivers improved performance over earlier methods by making fewer assumptions, fitting probes and SNPs across hybs, and borrowing information across SNPs
- Analogous to the time-proven and now industry-standard improvements seen between the newer RMA* and older MAS5 algorithms used in Affymetrix RNA gene expression analysis arrays
- In collaboration with and based on proof of concept work (RLMM) by Terry Speed lab. *(Rabbe and Speed, Bioinformatics, Nov. 2005)*
- Available as an open source, cross-platform tool, or as a graphical Windows application from www.affymetrix.com

* As well as other multi-chip probe-affinity-modeling methods (PLIER, GC-RMA, dChip, etc.)

AFFYMETRIX
The Way Ahead.™

Performance of new genotype calling algorithm

2 of our many validation data sets using HapMap DNA samples

		BRLMM at 0.3		BRLMM at 0.5		DM at 0.33	
		Call Rate	Concordance	Call Rate	Concordance	Call Rate	Concordance
Set1	All	96.60%	99.40%	98.40%	99.20%	95.00%	98.90%
	Hom	97.10%	99.40%	98.70%	99.30%	98.00%	99.40%
	Het	95.10%	99.40%	97.60%	99.10%	86.90%	97.40%
Set2	All	98.40%	99.50%	99.30%	99.40%	97.60%	99.30%
	Hom	98.10%	99.50%	99.20%	99.40%	98.70%	99.40%
	Het	99.00%	99.70%	99.60%	99.60%	94.60%	99.00%

- Conclusion: BRLMM eliminates majority of no-calls vs. DM
  - Higher call rates at better HapMap concordances
  - Heterozygote and homozygote call rates balanced
  - Best leverages the updated 500K assay protocol

AFFYMETRIX

The Way Ahead.™

Broad Institute Data

Nsp	**DM**	**BRLMM**
Samples (478 before filters)	478 (100%)	478 (100%)
SNPs (256,553 before filters)	199,282 (77.7%)	249,636 (97.3%)
Sample call rate	**97.40%**	**99.20%**
Mendel error rate (per genotype, per trio)	0.09%	0.10%

Stv	**DM**	**BRLMM**
Samples (739 before filters)	721 (97.5%)	726 (98.2%)
SNPs (233,477 before filters)	195,289 (83.6%)	225,035 (96.4%)
Sample call rate	**97.60%**	**99.10%**
Mendel error rate (per genotype, per trio)	0.15%	0.11%

"Based on significant improvements in every quality control metric examined, we see no reason to not immediately replace DM with BRLMM in all applications." -the Broad Institute

http://www.broad.mit.edu/gen_analysis/genotyping/brlmm_affy_ncrr.html

AFFYMETRIX
The Way Ahead.™

WGA-Amplified DNA on the Mapping 500K Arrays

Sample	Repli-g Call rate	Standard call rate	Reproducibility
1	97.3	98.6	99.91
2	95.9	96.8	99.82
3	95.3	96.7	99.76
4	96.6	96.7	99.84
5	97.2	95.6	99.81
6	95.3	96.0	99.75
7	97.4	95.7	99.88
8	95.8	98.4	99.84
9	93.7	96.4	99.67
10	94.9	96.9	99.77
11	95.8	96.8	99.84
12	95.1	97.4	99.71
Average	95.9	96.8	99.80

- Using 10 ng starting material
- Using Qiagen Repli-G

AFFYMETRIX®

The Way Ahead.™

Affymetrix Whole-Genome Strategy

Increasing the Power to Find Associations

AFFYMETRIX®

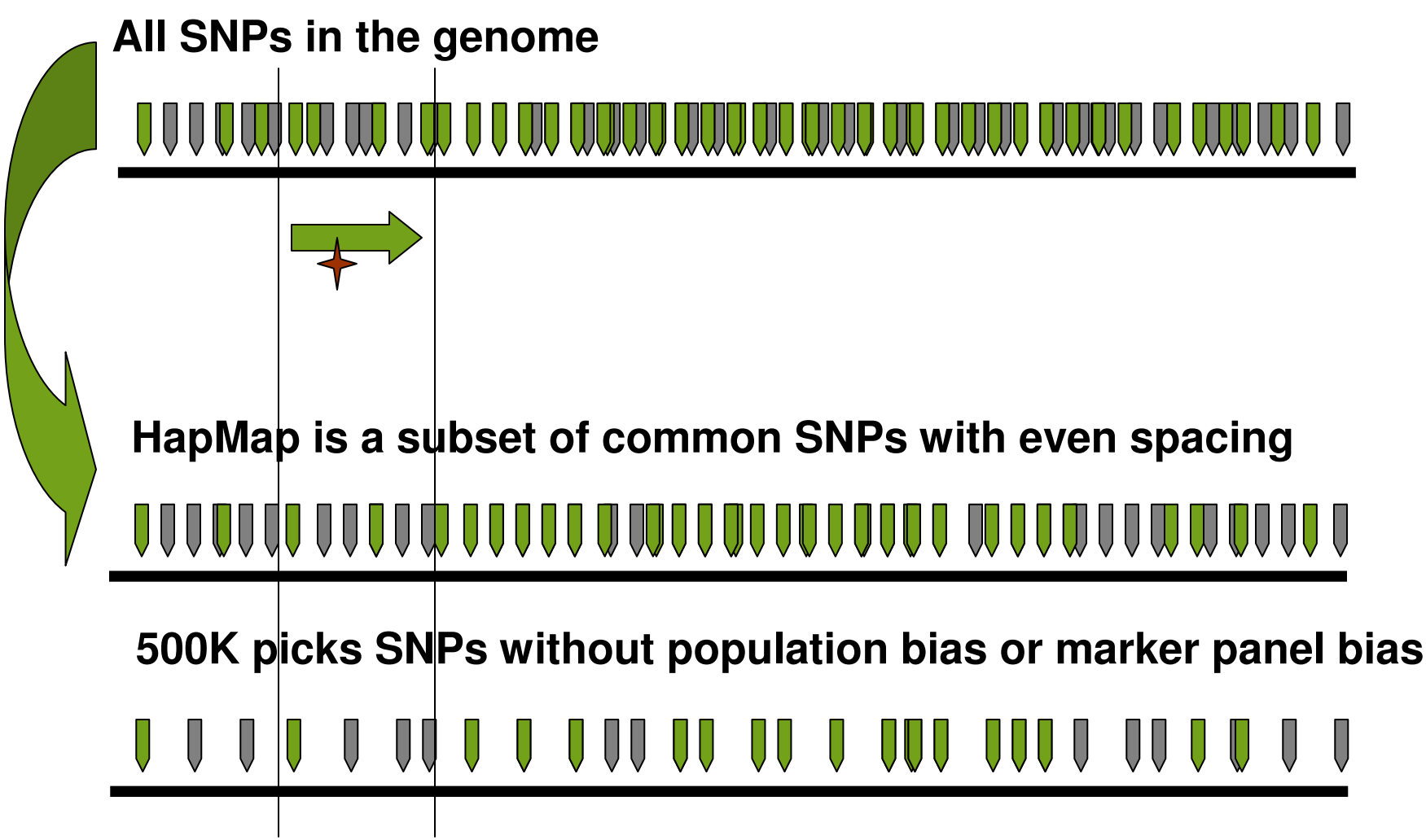
The Way Ahead.™

Genome Coverage of 500K SNPs

SNP Selection Strategy

AFFYMETRIX
The Way Ahead.™

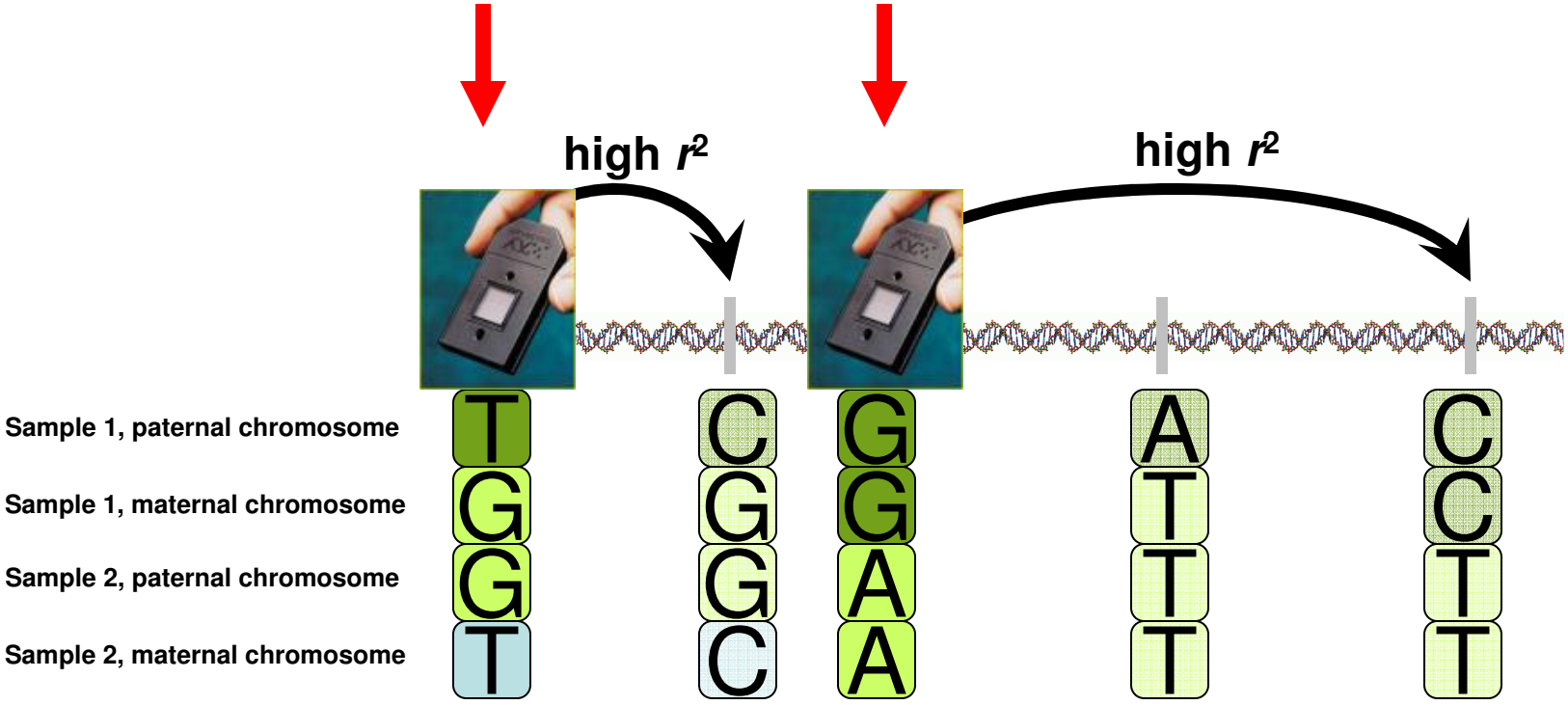
Genetic Coverage: Oversampling gives Robustness in Association Studies



AFFYMETRIX
The Way Ahead.™

Linkage Disequilibrium:
Typing a subset of SNPs captures many

Tests of association:



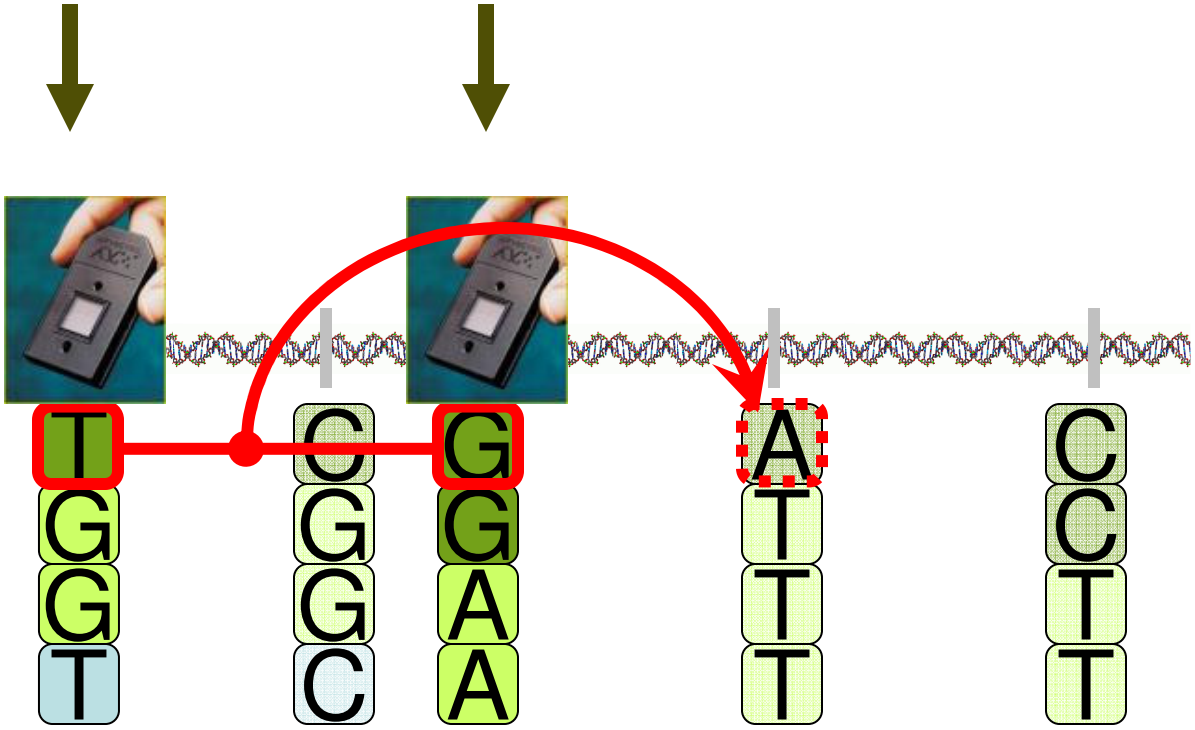
SNPs captured: ☑ ☑ ☑ ☑

Courtesy of Broad Institute

AFFYMETRIX
The Way Ahead.™

Capturing all of the information using Multimarker analysis

Tests of association:

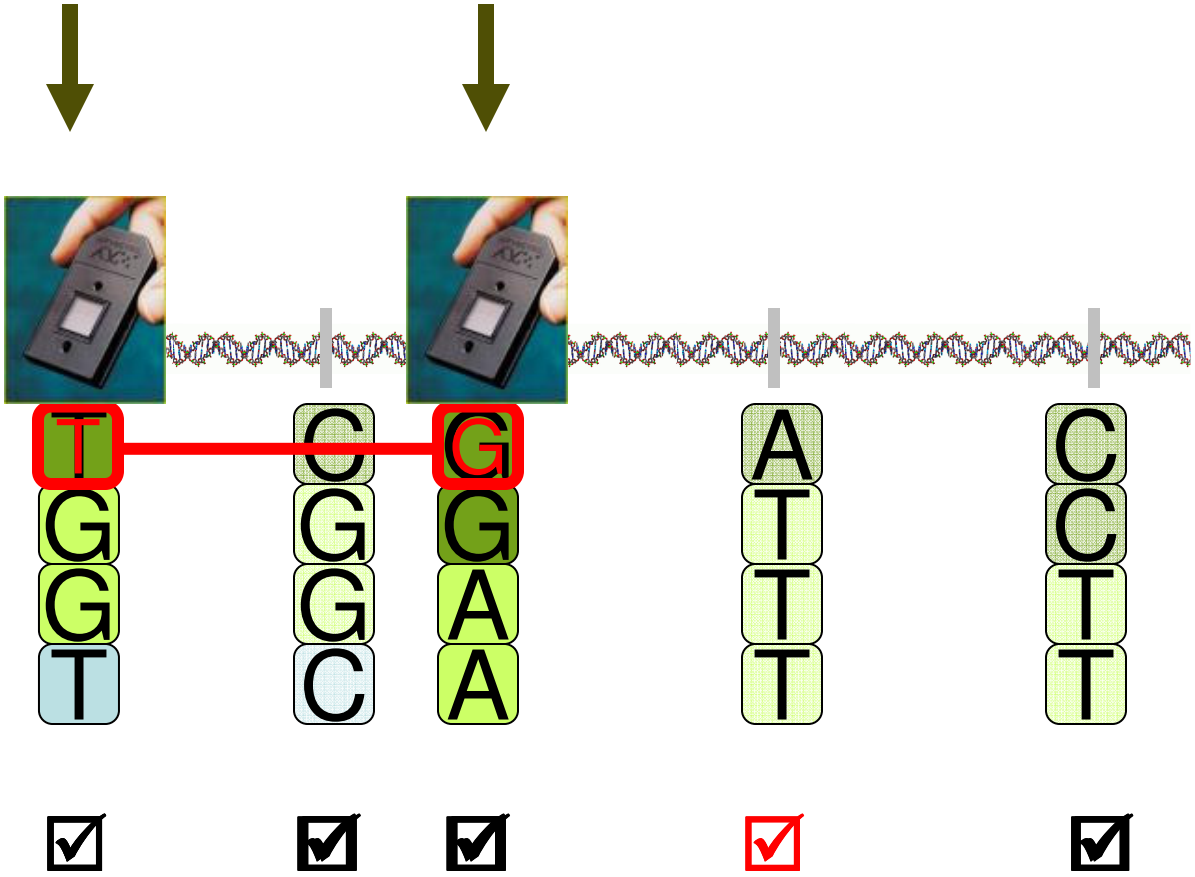


SNPs captured: ☑ ☑ ☑ ☑

Courtesy of Itsik Pe'er, Broad Institute

Capturing all of the information using Multimarker analysis

Tests of association:

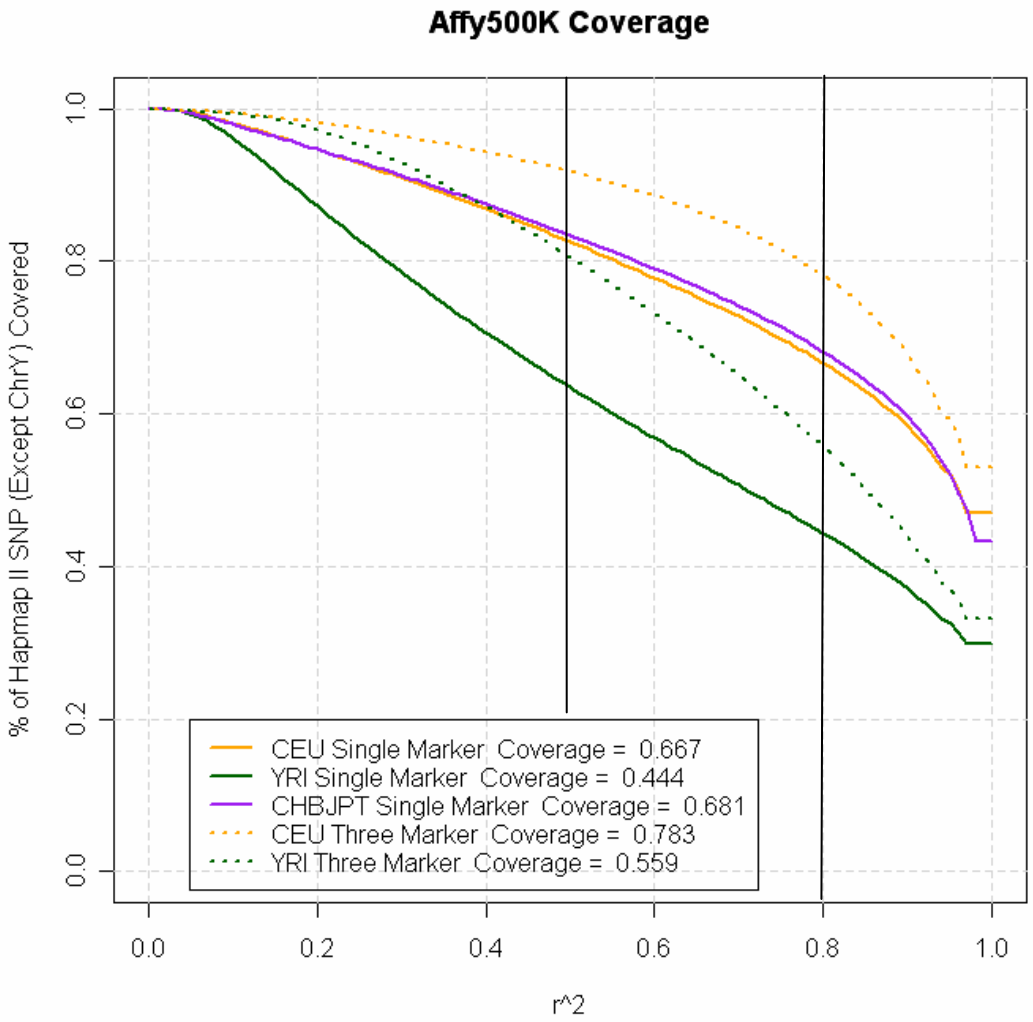


Note: Adds just one single pre-specified test

SNPs captured: ☑ ☑ ☑ ☑ ☑

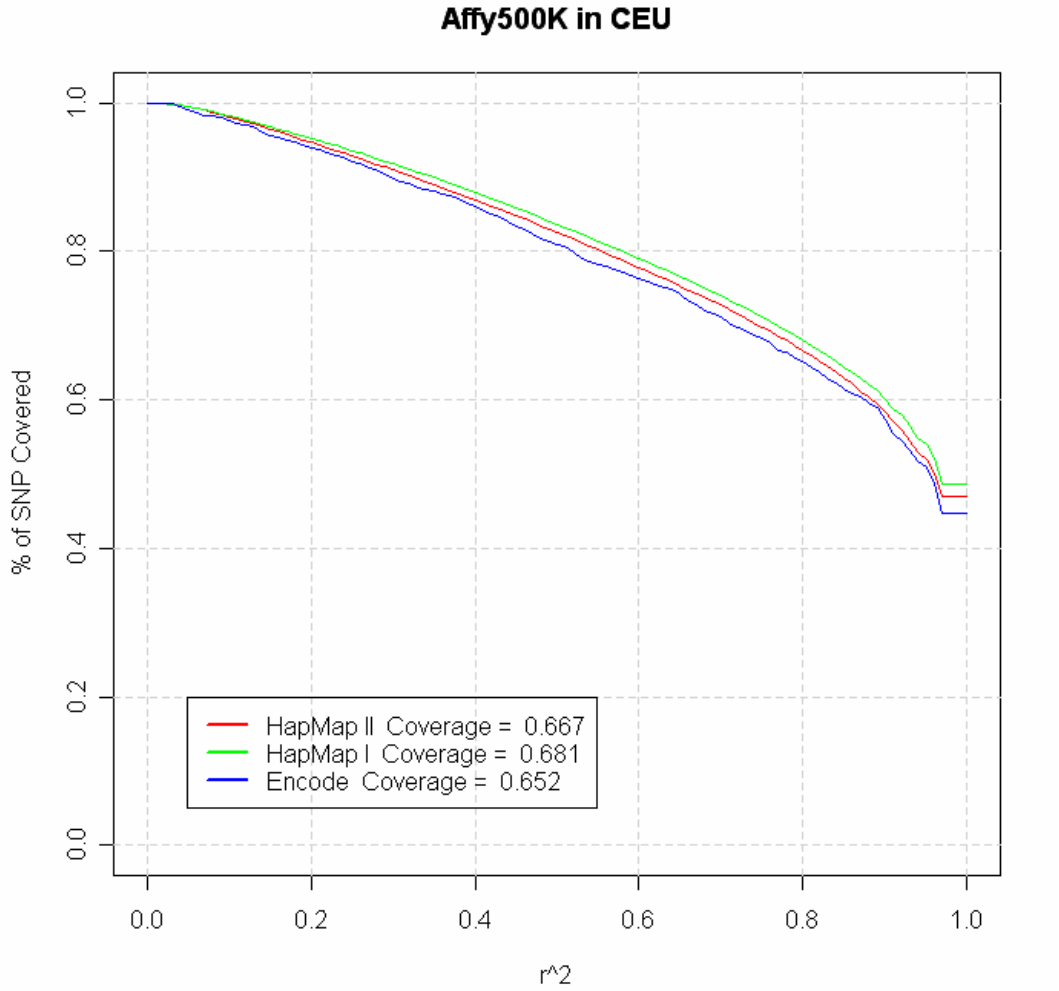
Courtesy of Itsik Pe'er, Broad Institute

Affymetrix Mapping 500K Coverage across Populations is consistent



- Mapping 500K shows consistent coverage across Asian and Caucasian Populations with single marker (pair wise) analysis
  - ~67% Coverage of CEU, CHB, JPT at r^2>0.8 using only pair-wise single marker analysis
- Multimarker analysis demonstrates coverage of almost 80% in Caucasian and Asian populations

Mapping 500K Has Consistent Coverage On HapMap I, HapMap II & ENCODE



- Mapping 500K SNPs were selected in an unbiased manner with respect to SNP list.
- Coverage (here in CEPH Utah population) is very similar when measured on different SNP lists

AFFYMETRIX
The Way Ahead.™

Things to bear in mind when considering coverage of SNP panels

- Pure Tag SNP panels take the reasonable approach of directly optimizing coverage using a particular population(s) and particular pool of SNPs.
  - Methods vary considerably, but are well-described in the literature
- However, remember that one **cannot** estimate coverage using the same dataset from which the SNP panel was designed
  - Yields an unrealistically inflated estimate of coverage in “real” studies
  - Can bias analyses if assumed as a prior

AFFYMETRIX
The Way Ahead.™

Things to bear in mind when using pure Tag-SNP panels

- Tag SNPs tend to transfer well, but they lose genetic coverage even when used in new similar populations:
  - From English to English (Zeggini et al, Nature Genetics, 2005)
  - From one Chinese sub-population to another (Huang et al, PNAS, 2005)
  - Between European Caucasian populations (Muller at al, AJHG 2005)
  - From CEPH to Finns (Willer et al, Genetic Epidemiology, 2006)
  - From Yorubans to African Americans (de Bakker et al, Pacific Symposium in Biocomputing, 2006)

AFFYMETRIX®
The Way Ahead.™

Oversampling Approach of Affymetrix Mapping 500K

- One can apply data filters without sacrificing coverage
- For ex: one can completely remove 15% of 500K SNPs with only a 3.5% drop in HapMap coverage
  - Multiple “chances” to detect association, resolve haplotypes
  - This is much less true of pure Tag-SNP panels

AFFYMETRIX
The Way Ahead.™

Effect of Missing Data Points

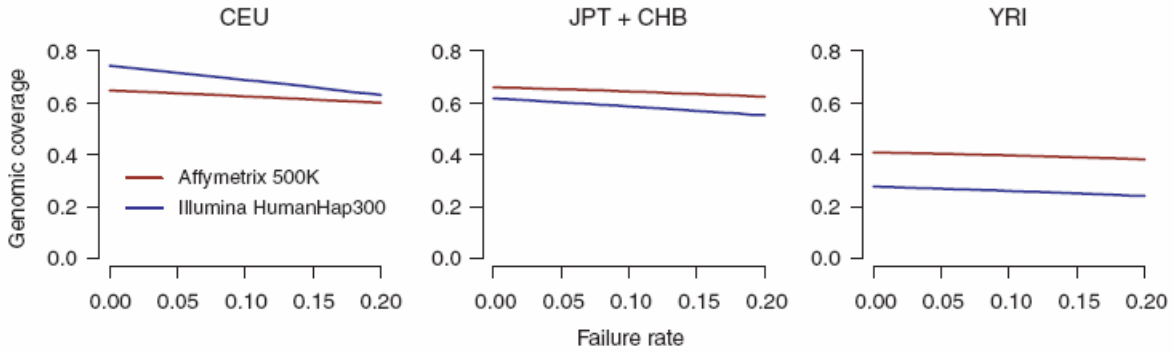


Figure 3 Coverage of common variation in the Phase II HapMap by the Affymetrix 500K and Illumina HumanHap300 products plotted as a function of random genotype failure rate. For each level of failure rate, markers were excluded at random, and coverage was calculated using only the remaining markers. Each point is the average of 1,000 replicates. The larger number of SNPs and increased redundancy in the Affymetrix array provide it greater resistance to decreased coverage due to marker failure. The Illumina array performs worse in non-CEU populations because its SNPs are so carefully targeted toward CEU.

Barrett and Cardon, Nature Genetics, 2006
Above graphs represent single marker, not multimarker analysis

AFFYMETRIX®

The Way Ahead.™

Affymetrix Whole-Genome Strategy

More Samples = More Power

AFFYMETRIX
The Way Ahead.™

Affymetrix Whole Genome Strategy

- Lower cost = more samples = higher genetic power
- Many studies are underpowered today
- Now you can afford to run the number of samples you need to for highly powered studies
- One chip pricing available today on Mapping 500K

AFFYMETRIX
The Way Ahead.™

GeneChip Mapping Product - Probe Array Tiling

-2	-1	0	+1	+2	+4
PMA	PMA	PMA	PMA	PMA	PMA
MMA	MMA	MMA	MMA	MMA	MMA
PMB	PMB	PMB	PMB	PMB	PMB
MMB	MMB	MMB	MMB	MMB	MMB

Quartet

- Multiple Quartets are evaluated per SNP

AFFYMETRIX
The Way Ahead.™

Increasing Price Performance

Feature Size	Number of Probes/SNP	Number of SNPs/array	Price/sample (array and reagents)
18um	40	**11,500**	**$490** $200 (10K 2.0)
8um	40	**60,000**	**$1000-$1400**
5um	24	**268,000**	**$500-$600**
5um	12	**500,000**	**$250**

AFFYMETRIX
The Way Ahead.™

Current Control Samples

All arrays have been purchased, with studies underway

Investigator	Institution	Population	# Samples
Schreiber	**Kiel - NGFN**	**POPGEN – N. German**	**800**
Meitinger	**GSF - NGFN**	**KORA-gen S. German**	**1800**
Lathrop	CNG	French	500
Maerz	Gratz	Luric – German	500
Arner	Karolinska	Swedish –	300
Uitterlinden	Erasmus	Dutch living in Rotterdam	500
Gether	Rigshospital	Danish	200
Lai/Nelson	**GSK**	**Mexican, European, Chinese, Japanese in first 1200**	5000
Offit	MSKCC	Ashkenazi Jewish	200
Donnelly et al	WTCCC	1958 Bristol Birth Cohort	3000

Total: 12,770

AFFYMETRIX®

The Way Ahead.™

New Products from Affymetrix for Association Studies

AFFYMETRIX

The Way Ahead.™

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

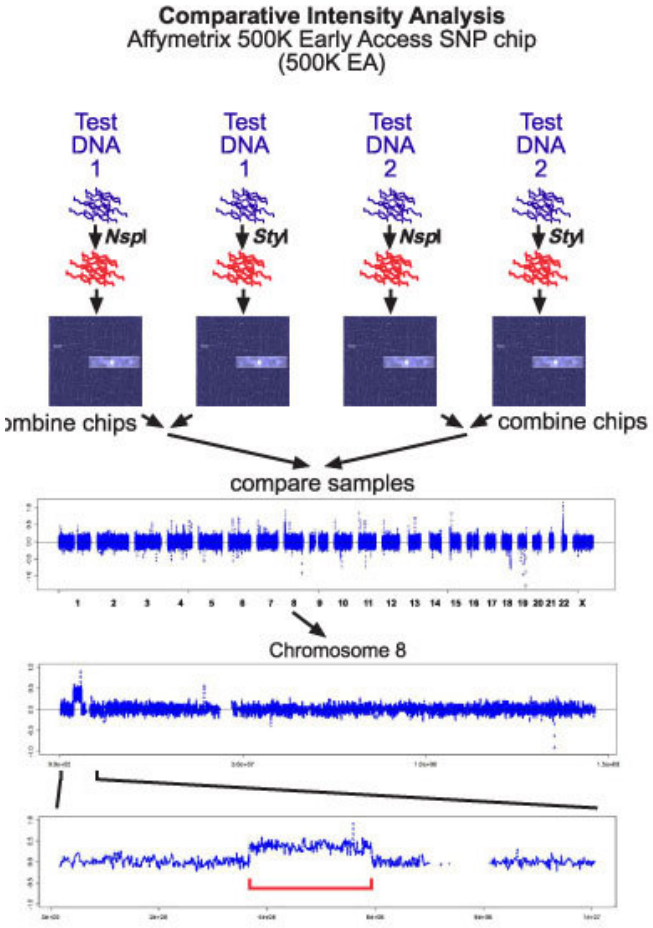
- Copy Number Variations (CNV's) are any regions in the genome of at least 1kb in size that demonstrate an amplification or deletion.
- Identified 1,447 CNV regions in 270 samples originally collected for the International HapMap Project.
- CNV regions average 250 kb in size encompassing some 12 percent of the genome in the samples tested and about 2,900 genes
- 285 of these genes are known to be associated with diseases based on their presence in the Online Mendelian Inheritance of Man database
- The Database of Genomic Variants has been created

Redon, R., et. al. Nature. 2006 Nov 23; 444: 444-454.

Komura, D., et. al. Genome Res. 2006. Advance online publication doi:10.1101/gr.5629106

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.™

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

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
  - 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

The Way Ahead.™