APPLICATION FOCUS:

ChIP-on-chip

Gene Regulation

Chromatin immunoprecipitation (ChIP) is an effective technique for enriching and identifying genomic DNA sequences bound by regulatory proteins, as well as identifying sites of histone and DNA modification. The full power of the application, however, is only realized when combined with very high-density, array-based detection. This allows interrogation of entire genomes at very high resolution and has already revealed a wealth of essential information about regulatory activity in non-coding and intergenic regions.

It is no longer sufficient, or necessary, to limit regulatory studies to promoter regions or defined genomic loci. An unbiased, whole-genome approach reveals the full regulatory network activity of transcription factors and epigenetic modifications. Affymetrix' unique high-density tiling arrays accommodate 6.4 million features per array, allowing whole-genome (human or mouse) coverage on just a few arrays, delivering high-performance, cost-effective whole-genome ChIP-on-chip analysis.

TABLE OF CONTENTS

ChIP Applications 2

The ChIP-on-chip Toolbox 5

Data Analysis Solutions 7

References 8

CRITICAL FEATURES

Why are researchers choosing Affymetrix products for whole-genome ChIP-on-chip analysis?

- Lowest number of arrays possible to cover the genome of interest
- High-resolution probe design to pinpoint multiple data points per binding region, resulting in high accuracy and sensitivity
- Easy-to-use tools for preliminary and advanced data analysis

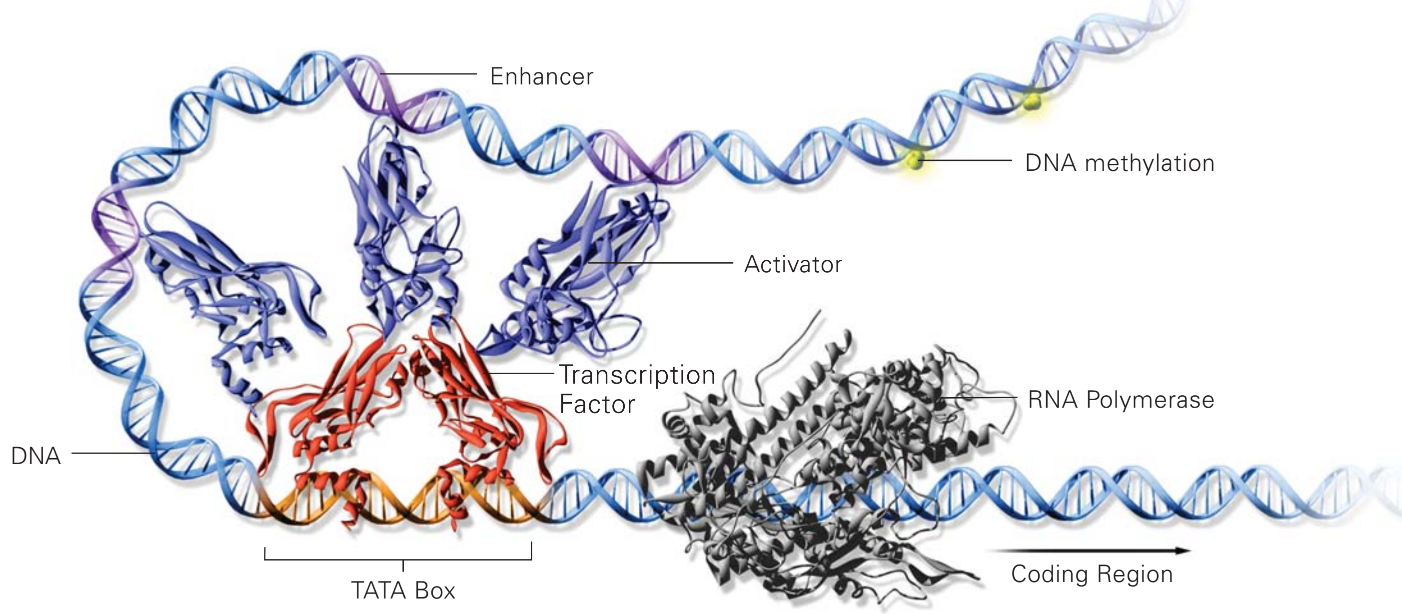


Figure 1: The total genome activity of transcription factors and epigenetic modifications are far more complex than previously predicted.

Only Affymetrix high-density arrays make whole-genome ChIP-on-chip experiments affordable and practical to conduct today—using just a single array for model organisms, and seven-array sets for human and mouse. These whole-genome products allow you to analyze all coding and non-coding regions for an unbiased view of the genome, which is especially important as non-coding regions contain several enhancer/repressor binding sequences. Single arrays from mouse and human whole-genome sets are also available for your initial assay development and optimization. Your initial experiments give a realistic chromosome-wide view of activity, and can be easily scaled up for whole-genome analysis.

Recent reports have demonstrated that epigenetic and transcription regulatory networks are far more complex than previously predicted, with individual transcription factors binding to thousands of sites across the genome, and with much of the regulatory activity located at distal sites far from promoter regions [5, 6].

Previous ChIP-on-chip experiments using focused arrays (as opposed to whole- genome arrays) have traditionally analyzed candidate genes, annotated promoter regions or customized loci. These approaches have revealed only a partial picture of the total genome activity of key transcription factors and epigenetic modifications.

With 25-mer probes spaced every 35 base pairs (bp) across an entire genome, Affymetrix Tiling Arrays offer complete and unbiased genome coverage on a very small number of arrays. They offer the only practical, cost-effective and scalable approach to gene regulation analysis available to date.

Start your analysis with a single array from a set, and confidently scale up to the entire genome, using the seven-array set. No need to repeat large amounts of work—whole-genome analysis is affordable and practical now.

Affymetrix whole-genome arrays have already been used to map sites for transcription factor binding, and DNA and histone modification, revealing key information about the functions of various transcription regulators (including estrogen receptor [ER] and p53), and helping identify target genes involved in disease progression and development—insights that could not have been made using focused arrays.

Affymetrix Offers Unique Off-the-shelf Tools for Advanced Gene Regulation Analysis, Which Provide Unrivalled Genome Coverage

- **Content:** Over 6.4 million features per array for human and mouse sets
- **Simplicity:** Low number of arrays for human and mouse whole genomes, single whole-genome arrays for other organisms
- **Resolution:** 35 base pairs from midpoint to midpoint of adjacent oligos
- **Specificity:** 25-mer probes deliver high sensitivity with a low incidence of false positives in initial tests

APPLICATIONS FOR IP-BASED ANALYSIS OF GENE REGULATION

The basics of Chromatin IP make it applicable to the study of any protein interaction with genomic DNA, or DNA or histone modification, provided an antibody exists that is specific for that protein or modification, and is suitable for immunoprecipitation.

Protein-DNA complexes are chemically cross-linked, then purified from the cell and immunoprecipitated using an antibody specific for the regulatory protein or modification under scrutiny. The DNA in the enriched fraction can be amplified using real-time PCR to confirm the presence of expected sequences, and then applied to arrays for comparison with non-enriched samples for identification of all binding sequences—known and novel.

The ChIP-on-chip assay can be used to examine gene regulation by both epigenetic modifications, such as histone and/or DNA modification, and by transcription machinery, such as transcrip-

tion factor binding. ChIP-on-chip has also been used to study other protein-mediated processes such as DNA replication and repair.

Mapping Transcription Factor Binding Sites

Recent evidence shows that many transcription protein-binding sites in the genome are actually in non-coding regions once considered "junk" DNA, and can sometimes be up to 100 kilobases (kb) from the gene they regulate. In addition, transcription factors are more active than previously thought, with a high number of binding sites, not always predictable from either canonical promoter regions or "consensus motif" searches.

Carroll, *et al.* (2006) published a genome-wide map of estrogen receptor and RNA Polymerase II binding sites in breast cancer cells. The study identified cis-binding sites in previously unexplored regions of the genome, and investigated cooperation between transcription factors in estrogen signaling associated with breast cancer. Q-PCR validation of 15 randomly selected binding sites confirmed ER recruitment to all of the ChIP-on-chip identified sites and none of the negative controls. Dr. Myles Brown, an author of the paper, discussed this work in the October 2005 issue of the *Affymetrix Microarray Bulletin* (AMB), a magazine that highlights key microarray-based research from around the world.

Cawley, *et al.* (2004) used Affymetrix Tiling Arrays to map binding sites for the transcription factors Sp1, cMyc and p53. Thousands of previously unknown transcription factor-binding sites were identified—only about 22 percent of which were located at the 5′ end of protein-coding genes—disproving common assumptions about the location of transcription factor-binding sites. Dr. Kevin Struhl, an author of the paper, discussed this work in the October 2005 issue of the *Affymetrix Microarray Bulletin.*

Using a human whole-genome set from Affymetrix, Yang, *et al.* (2006) examined the global binding behavior of p63 and the evolutionary conservation of p63 binding sites.

Mapping Histone Modifications

Bernstein, *et al*. (2005) used Affymetrix Tiling Arrays to map the location of two methylated lysines in the histones of embryonic stem cells. This modification pattern appears to be required to silence developmental genes while keeping the cells poised to differentiate. An interview with Dr. Bernstein discussing this work can be found in the October 2006 issue of the *Affymetrix Microarray Bulletin*.

Bernstein, *et al.* previously used tiling arrays to map histone methylation and acetylation across non-repetitive portions of human chromosome 21 and 22 in both the human and mouse genomes. Bernstein's team determined that sites of trimethylation correlated with transcription start sites and that methylation sites are strongly conserved between the two species.

Mapping DNA Methylation

Zhang, *et al.* (2006) recently reported the first genome-wide analysis of DNA methylation in *Arabidopsis thaliana* at 35-bp resolution, using Affymetrix Tiling Arrays. Using an anti-methylcytosine antibody to precipitate methylated fragments of the genome, the study found that 19 percent of the Arabidopsis genome is methylated. They unexpectedly found that 35 percent of expressed genes contain some methylation within coding regions, while around 5 percent show methylation in promoter regions.

"The combination of this unique resource [whole-genome ChIP-on-chip analysis at 35-bp resolution] with gene expression data sets serves to elucidate the mechanisms underlying estrogen-regulated gene expression in breast cancer." Carroll, *et al.* (2006)

AFFYMETRIX
MICROARRAY
BULLETIN

Thousands of Previously Unknown Transcription Factor Binding Sites Mapped Using Tiling Microarrays

Harvard's Kevin Struhl and Dana-Farber's Myles Brown discuss ChIP-on-chip as an unbiased method for transcription factor binding site identification

Dr. Myles Brown in the October 2005 edition of the *Affymetrix Microarray Bulletin.*

Distribution of All TFBS Regions

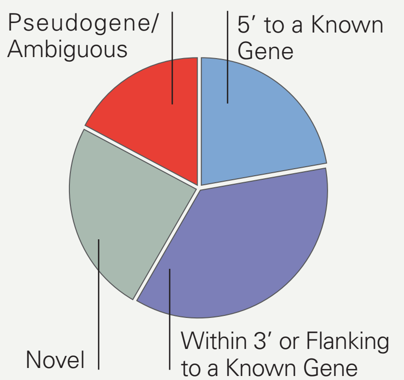


Figure 2: Cawley, *et al.* (2004) Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell* 116, 499-509.

Hayashi, *et al.* (2006) employed a similar approach to map DNA methylation in the ENCODE regions of the human genome, identifying over 700 candidate methylation sites. They also compared DNA methylation with histone H3 and H4 acetylation, finding a reciprocal relationship between the two.

Genpathway, an Affymetrix-approved service provider, has developed a DNA immunoprecipitation (IP) assay to enrich and detect methylated sequences on Affymetrix Human Whole-genome Tiling Arrays. Genpathway recently analyzed DNA methylation in MDA-MD-231 cells, an estrogen receptor-negative breast cancer cell line (http://genpathway.com/). On a single array covering three chromosomes (8, 11 and 12), they observed 997 methylation sites, of which only 223 were in or close to CpG islands (see Table 1), strengthening the argument for a whole-genome, unbiased approach to epigenetic mapping using Affymetrix whole-genome arrays. In addition, Genpathway's IP approach using its optimized protocols results in the identification of methylated DNA regions with higher sensitivity and specificity than other methods.

Genpathway offers a uniquely comprehensive package of chromatin-IP or methylated DNA-IP assay design and validation, array processing (using Affymetrix high-density, whole-genome Tiling Arrays) and advanced data analysis. An interview with Dr. Mary Warren, Chief Scientific Officer of Genpathway, discussing these services can be found on the Affymetrix *UserForum*. These services include a wide range of ChIP assays covering gene expression and gene regulation on a range of genomes. For further information, visit www.genpathway.com or contact your local Affymetrix representative.

Mapping Other Protein-mediated Biological Pathways

Lengronne, *et al.* (2004) used high-density tiling arrays to study how transcriptional activity determines where sister chromosomes attach during cell division, an important process in chromosome segregation that goes awry in cancer. Cohesin, the protein that holds sister chromatids together, binds genomic DNA at unspecified sites, and the unbiased and high-resolution ChIP-on-chip approach showed a striking localization of cohesin to points of converging transcription. Dr. Frank Uhlmann, principal investigator on the study, discussed the group's findings in the October 2005 issue of the *Affymetrix Microarray Bulletin*.

Table 1: Locations of Peaks Representing Methylated DNA

Peaks on Array F (Chromosomes 8, 11, 12)	Number	Percent
Total number on Array F	997	
Extrapolated to entire genome	6070	
In CpG Islands +/- 200 bp	233	23.4%
In Promoters (-1 to -10,000 bp from TSS*)	123	12.3%
In genes	584	58.6%
In exons +/- 100 bp	137	13.7%
Unannotated (no gene within 10 kb)	216	22.0%

*TSS = Transcription Start Site1

Data provided by Genpathway

"Our method recognizes DNA methylation with little bias for genomic location and therefore is useful for comprehensive, high-resolution analysis of DNA methylation." Hayashi, *et al.* (2006)

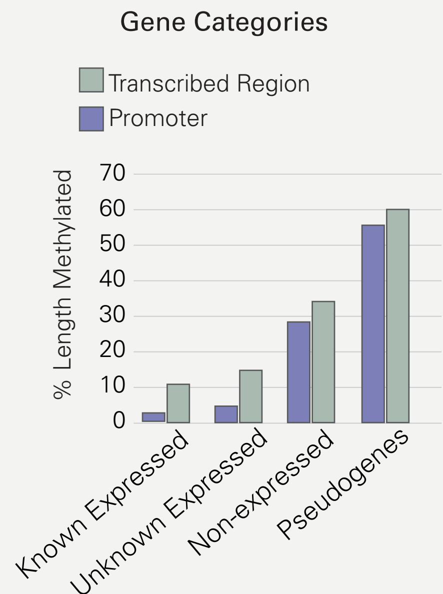


Figure 3: Zhang X, *et al.* (2006) Genome-wide High-resolution Mapping and Functional Analysis of DNA Methylation in *Arabidopsis*. *Cell* 126, 1-13.

"It will be important to examine the entire genome in ES cells as well as to follow their fate during differentiation and development." Bernstein, *et al.* (2005)

THE AFFYMETRIX CHIP-ON-CHIP TOOLBOX

The Affymetrix ChIP-on-chip toobox is designed to help you through every stage of the ChIP-on-chip workflow. The ChIP-on-chip toolbox includes assay design and validation guidelines, reagents for sample labeling and preparation, arrays for ChIP-on-chip, data analysis solutions, and community support. This toolbox includes the Affymetrix ChIP-on-chip protocol, which serves as a "start-to-finish" guidance document on experimental design for a ChIP-on-chip assay.

ChIP-on-chip is a powerful technique for identifying sites of genomic regulation, but can be a challenging task to undertake for the first time. Affymetrix provides the tools necessary to perform the assay, and offers advice on assay optimization and validation.

Assay Design and Validation Guidlines

The Affymetrix ChIP-on-chip protocol is available for download from www.affymetrix.com. However, ChIP-on-chip is a complex process, and each step of the protocol will need to be validated by users in their own applications. Particular areas for validation are:

- Verification of antibody utility for ChIP (specificity, enrichment factor, etc.)
- Crosslinking and sonication conditions
- Availability of known positive control for IP enrichment QC
- What negative controls will work best-Non-specific IP/Mock IP/Input?
- Amplification and fragmentation of post-IP DNA
- QC for enrichment pre- and post-amplification

Reagents for Sample Labeling and Preparation

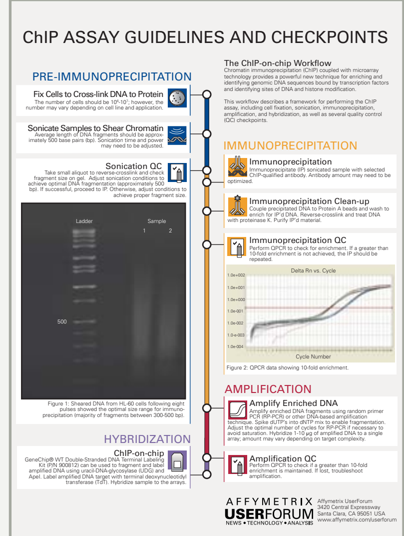
Affymetrix provides all of the reagents you need to label, purify and hybridize your sample post-amplification. Reagents for ChIP and amplification steps are user-supplied.

Table 2: Reagents for ChIP-on-chip Applications	
Reagent	Product Code
GeneChip® WT Double Stranded DNA Terminal Labelling Kit - 30 reactions	900812
GeneChip® Sample CleanUp Module - 30 reactions	900371
GeneChip® Hybridization, Wash and Stain Kit - 30 reactions	900720
GeneChip® Control OligoB2, 3nM - 30 reactions	900301

The Affymetrix ChIP-on-chip Toolbox

Designed to help you through every stage of the ChIP-on-chip workflow.

- Assay design and validation guidelines
- Reagents for sample labeling and preparation
- Arrays for ChIP-on-chip
- Data analysis solutions
- ChIP-on-chip community support
- The protocol is available for download from www.affymetrix.com/support



The Affymetrix ChIP-on-chip workflow guides you through the key steps of the ChIP-on-chip assay. This workflow is available on the Affymetrix *UserForum* at: www.affymetrix.com/userForum/index.uf

Whole-genome Arrays for ChIP-on-chip

Affymetrix Whole-genome Tiling Arrays for ChIP-on-chip analysis are available for human, mouse, Arabidopsis, *C. elegans*, *S. cerevisiae*, *S. pombe* and Drosophila.

Table 3: Whole-genome Tiling Arrays for ChIP-on-chip

Array	Product Code	Pack Size
Human Tiling 2.0R Array Set (7 array set)	900772	1
Mouse 2.0R Array Set (7 array set)	900852	1
Arabidopsis Tiling 1.0R Array	900594	6
S. cerevisiae Tiling 1.0R Array	900645	6
S. pombe Tiling 1.0FR Array	900647	6
C. elegans Tiling 1.0R Array	900935	6
Drosophila Tiling 1.0R Array	900588	6

Table 4: Single Arrays from Human Whole-genome Set

GeneChip® Array	Product Code	Pack Size
Human Tiling 2.0R Array A (chr 1, 6)	900779	6
Human Tiling 2.0R Array B (chr 2, 9, 19)	900780	6
Human Tiling 2.0R Array C (chr 23, 22, 21, X,Y,and mitochondria)	900781	6
Human Tiling 2.0R Array D (chr 4, 5, 18, 20)	900782	6
Human Tiling 2.0R Array E (chr 5,7, 16)	900783	6
Human Tiling 2.0R Array F (chr 8, 11, 12)	900784	6
Human Tiling 2.0R Array G (chr 10, 13, 14, 17)	900785	6

Table 5: Single Arrays from the Mouse Genome Set

GeneChip® Array	Product Code	Pack Size
Mouse Tiling 2.0R Array A (chr 1, 9, 19)	900984	6
Human Tiling 2.0R Array B (chr 2, mitochondria, X, Y and unassigned)	900781	6
Mouse Tiling 2.0R Array C (chr 3, 7, 18)	900897	6
Mouse Tiling 2.0 R Array D (chr 4, 11, 17)	900895	6
Mouse Tiling 2.0 R Array E (chr 2, 12, 15)	900898	6
Mouse Tiling 2.0R Array F (chr 6, 8, 16)	900899	6
Mouse Tiling 2.0R Array G (chr 10, 13, 14)	900900	6

Focused and Custom Arrays for ChIP-on-chip

For research with a specific interest in promoter regions or ENCODE regions, for example, focused arrays are available. Custom Arrays are also often an attractive approach to ChIP-on-chip experiments. You can define any content, at any resolution, any strand detection, any application with a range of array formats accommodating from 38,000 to 6.5 million unique oligonucleotides per array.

Table 6: High-density Tiled Arrays for Promoter and ENCODE Regions

GeneChip® Array	Product Code	Pack Size
Human Promoter 1.0R Array	900775 900776 900777	2 6 30
Human Chromosomes 21/22 2.0R Array	900936	6
ENCODE 2.0R Array	900937	6
Mouse Promoter 1.0R Array	900889 900890 900891	2 6 30

Data Analysis Solutions

Affymetrix offers its own tools for preliminary data analysis and visualization of peaks, and can be freely downloaded from www.affymetrix.com/support.

- Tiling Analysis Software (TAS) will perform basic raw data analysis to produce signal and *p*-values for each genomic probe on the array.
- Integrated Genome Browser (IGB) allows visualization of data taken from TAS. You can select and load genomic annotations for cross-reference with ChIP-on-chip peak data.

There are, however, a growing number of other commercial and "freeware" packages to help you with ChIP-on-chip data analysis. These include:

The MAT Algorithm from the Dana-Farber Cancer Institute

A novel analysis algorithm to reliably detect regions enriched by ChIP on Affymetrix tiling arrays developed by the laboratory of Dr. Xiaole Shirley Liu.

- Can detect enriched regions from individual ChIP samples
- Robust *p*-value calculation and false discovery rate estimate
- Eliminates the need for normalization
- Freely available at: http://chip.dfci.harvard.edu/~wli/MAT/

Partek® Genomics Suite

Partek® Genomics Suite™ integrates advanced statistics and interactive visualization to reliably extract biological signals from noisy data. Designed for high-dimensional genomic studies, Partek GS is fast, memory efficient and will analyze large data sets on a personal computer. It supports a complete, easy-to-use workflow and is integrated with public genomic resources including GenBank®, NCBI GEO and NetAffx™.

- Powerful statistics and interactive visualization
- Fast and memory-efficient
- Easy-to-use workflows designed for scientists
- Identify regions of protein/DNA binding in ChIP-on-chip studies
- Import and estimate mapping sites of protein/DNA interaction
- Detect, display and report mapping sites of protein/DNA interaction
- Use both reference and user samples as baseline
- Create lists of genes and SNPs in binding regions
- Remove technical and biological batch effects
- www.partek.com

Genomatix ChipInspector

A powerful program that extracts information from the expression level of single probes using Affymetrix tiling arrays.

- Increases the number of significant features while reducing the rate of false positives
- Utilizes the world's largest database of alternative transcripts
- Assigns probes correctly to transcripts and genes
- Eliminates interpolation/normalization problems
- Available at www.genomatix.de

REFERENCES

1. Bernstein B., *et al.* A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cell Development. *Cell* **125**:315-326 (2006).

2. Bernstein B., *et al.* Genomic Maps and Comparative Analysis of Histone Modification in Human and Mouse. *Cell* **120**:169-181 (2005).

3. Carroll J., *et al.* Genome-Wide analysis of Estrogen Receptor Binding Sites. *Nature Genetics* **38**(11):1289-97 (2006).

4. Carroll J., *et al.* Chromosome-Wide Mapping of Estrogen Receptor Binding Reveals Long-Range Regulation Requiring the Forkhead Protein FoxA1. *Cell* **122**:33-43 (2005).

5. Cawley S., *et al.* Unbiased Mapping of Transcription Factor Binding Sites Along Human Chromosomes 21 and 22 Points to Widespread Regulation of Noncoding RNAs. *Cell* **116**:499-509 (2004).

6. Euskirchen G., *et al.* CREB Binds to Multiple Loci on Human Chromosome 22. Molecular and *Cellular Biology* **24**(9):3804-14 (2004).

7. Hayashi H., *et al.* High-resolution DNA Methylation Mapping in Human Genome Using Oligonucelotide Tiling Array. *Human Genetics* **120**:701-11 (2007).

8. Lengronne A., *et al.* Cohesin Relocation from Sites of Chromosomal Loading to Places of Convergent Transcription. *Nature* **430**(6999):573-8 (2004).

9. Yang A., *et al.* Relationships between p63 Binding, DNA Sequence, Transcription Activity, and Biological Function in Human Cells. *Molecular Cell* **24**:593-602 (2006).

10. Zhang X., *et al.* Genome-wide High-Resolution Mapping and Functional Analysis of DNA Methylation in Arabidopsis. *Cell* **126**:1-13 (2006).

AFFYMETRIX Inc. USA
3420 Central Expressway
Santa Clara, CA 95051 USA
Tel: 408-731-5000 (1-888-DNA-CHIP)
Fax: 408-731-5380
www.affymetrix.com

AFFYMETRIX, UK Ltd.
Voyager, Mercury Park,
Wycombe Lane, Wooburn Green
High Wycombe HP10 0HH
United Kingdom
Tel: +44 (0) 1628 552550
Fax: +44 (0) 1628 552585
saleseurope@affymetrix.com
supporteurope@affymetrix.com

AFFYMETRIX JAPAN K.K.
Affymetrix Japan K.K.
Mita NN Bldg 16 Floor
4-1-23 Shiba
Minato-ku, Tokyo 108-0014
Tel: *81-(0)3-5730-8100
Fax: *81-(0)3-5730-8101
salesjapan@affymetrix.com
supportjapan@affymetrix.com
www.affymetrix.com

Please visit our web site at www.affymetrix.com for international distributor contact information.

For research use only. Not for use in diagnostic procedures. Part No.702410-1

©2007 Affymetrix, Inc. All rights reserved.

Affymetrix® ™, , ™, GeneChip®, HuSNP®, GenFlex®, Flying Objective™, CustomExpress®, CustomSeq®, NetAffx™, Tools To Take You As Far As Your Vision®, The Way Ahead™, Powered by Affymetrix™, GeneChip-compatible™, and Command Console™ are trademarks of Affymetrix, Inc. All other trademarks are the property of their respective owners. Array products may be covered by one or more of the following patents and/or sold under license from Oxford Gene Technology: U.S. Patent Nos. 5,445,934; 5,700,637; 5,744,305; 5,945,334; 6,054,270; 6,140,044; 6,261,776; 6,291,183; 6,346,413; 6,399,365; 6,420,169; 6,551,817; 6,610,482; 6,733,977; and EP 619 321; 373 203 and other U.S. or foreign patents.