GENE MAPPING IN COMPLEX TRAITS: GENE EXPRESSION PROFILING AND COPY NUMBER VARIABILITY IN GENE MAPPING OF AGE-RELATED COGNITIVE DECLINE

By

Beth Wilmot, MS

A DISSERTATION

Presented to the Department of Molecular and Medical Genetics

And the Oregon Health & Science University

School of Medicine

in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

August 2007

School of Medicine
Oregon Health & Science University

CERTIFICATE OF APPROVAL

This is to certify that the Ph.D. dissertation of

Beth Wilmot, MS

has been approved

Patricia L. Kramer, PhD
Co-Mentor/Advisor

Shannon K. McWeeney, PhD
Co-Mentor/Advisor

Jeffrey A. Kaye, MD
Member

Susan B. Olson, PhD
Member

John K. Belknap, PhD
Member

Table of Contents

Acknowledgements

My graduate work would not have been possible without the help of a number of people. I would like to give my heartfelt thanks to these people in recognition for all of their assistance. The members of my graduate committee who have been with me for five years Patricia Kramer, Shannon McWeeney, Jeffery Kaye, and Susan Olson. Many thanks go to Patti for supporting me and my research in many ways throughout these years as my mentor, for giving me research space and sending to me to multiple workshops and conferences. Special thanks to Patti for supporting me in doing the research I wanted. Many thanks go to Shannon for agreeing to be my co-mentor and especially for the enormous amount of her time and inexhaustible patience in helping me with my desire to learn algorithms and data analysis. Special thanks go to Tomi Mori and Shannon for giving me office space in their group and time on the fast computer, Einstein. I appreciate all of the assistance from Jeff and the Layton Aging and Alzheimer's Disease Center for supporting me and my research throughout my work, both through including me as a working member of the Center and giving me laboratory space. Susan Olson has been invaluable as a resource both as a medical geneticist and for her patience and wisdom as a graduate student advisor. I thank John Belknap for graciously agreeing to chair my exam committee. I thank all of my committee members for their advice and unwavering belief in me as a promising scientist.

Many people helped me in my research. From the Layton Aging and Alzheimer's Disease Center, Robin Guariglia helped me navigate the vast database in order to collect subjects for my studies. Jamie Laut, the technician in the Genetics Core, both put up with me while I trained her and then completed all of the qRT-PCR, RNA and DNA isolation

for my samples. Randy Nixon showed me how to read the neuropathology of Alzheimer's disease on tissue sections and gave me much scientific advice. David Clark in Randy's lab showed me the tissue section stains and helped me with collecting all of the slides for my review.

The Affymetrix Microarray Core performed all of the hybridizations to the gene expression arrays and genotyping arrays. My thanks go to the entire lab for their cheerful help in coping with my samples. Special thanks go to Chris Harrington and Kristina Vartanian for their advice and steadfast support throughout my projects.

Veena Rajaraman of the Cancer Center Informatics Shared Resource BioDevelopment Core showed me how to write for loops in R and gave me a great deal of encouragement while I was writing my genotyping scripts. Ted Laderas was always available for an S.O.S. when I couldn't get a script working properly. Ranjani Ramakrishnan helped me create my CNV database and helped me with my SQL queries. Brinda Selvaraj, Sharon Pearce and Eilis Boudreau have given me a great deal of encouragement.

My rotation advisors, Tim Stout and Cheryl Maslen were great role models and I enjoyed my time in their laboratories. Binoy Appukutan has been invaluable help in all aspects of my studies. He and Trevor MacFarland of Tim Stout's lab helped with many discussions of molecular biology and genetics. Brenda Polster and Shawn Westaway were always available for discussions. Betsy Ferguson has given me much encouragement and advice as well as many interesting scientific discussions. I am especially indebted to her for telling me about the "5th year hump".

The staff of the MMG department has been of enormous assistance, especially

Michele Neuhaus, Autumn Linde and Glenda Barton who have cheerfully helped me stay on track, encouraged me and given me good advice.

Many people not directly involved with my research at OHSU gave me an enormous amount of support and encouragement throughout these last years. My neighbors, who are always interested in my research and helped me with my house and yard, and especially Marcy Brillman for being a marathon training partner. In addition, my dear friends Binoy Appukutan, Ranjani Ramakrishnan, Lynn Rust and Joel Peterson who have made sure I did many fun and exciting things other than graduate school. In particular, I thank Susan Dalton without whose invaluable help I would not have been able to accomplish much of what I did.

I thank my family, Frank and Lou Wilmot, Denise Thomas and Wanda Anderson who have put up with a great deal and supported me emotionally and financially through many trials. I dedicate my dissertation to my parents who have done as much as parents could possibly do to support me in everything I’ve done.

I would like to thank the following groups for their financial support to me through Patti and the Layton Aging and Alzheimer’s Disease Center: the NIH/NIA P30-AG00817, NIH R01-NSO26656, the Murdock Charitable Trust, and the Medical Research Foundation of Oregon.

Abstract

Alternative strategies for mapping genes involved in common traits are needed due to the complexity of the underlying genetic and environmental causes. Improvements in technology have provided the ability to genotype increasing numbers of single nucleotide polymorphisms (SNPs) in multiple individuals in a cost effective manner. The advent of whole genome expression profiling technology has enabled investigators to combine functional genomics with genetic variation such as SNPs and copy number polymorphisms to inform the choice of possible candidate genes for further study

In this study, genes involved in age-related cognitive decline as defined by Alzheimer's disease (AD) were identified through the use of current high-throughput technologies. Gene profiling was used to identify functional candidates and associated pathways implicated in neuropathological phenotypes related to cognitive decline. Whole genome association (WGA) techniques were investigated to assess both the low level characteristics of the genotyping algorithms and the impact of DNA structural variation on the previous gene expression results.

Chapter I. Introduction

A. Overview

Historically, most advances in gene mapping have occurred through the use of linkage analysis in family studies of Mendelian diseases [1]. Translating these successes into finding genes involved in complex traits has been problematic. The main challenges are that complex traits are comprised of multiple genes of varying effect sizes and that epistatic interactions among multiple genes and the environment contribute to the phenotype [1-4]. Studies testing for association of genetic variations and phenotype in unrelated individuals may have greater power to identify genes of lower effect size than linkage analysis, but the sample sizes required to obtain significance across multiple genomic markers made these studies cost prohibitive for many years [5, 6]. As improvements in technology provided the ability to genotype increasing numbers of single nucleotide polymorphisms (SNPs) in multiple individuals in a cost effective manner, association studies proliferated. Despite a rise in the number of studies performed, few genes were replicated in other studies and therefore could not be identified unequivocally [3 , 7]. The lack of reproducibility of gene identification in complex traits is related to the low power of most association studies due to phenotypic heterogeneity and population heterogeneity, lack of controls for environmental factors and the only recent appreciation of the extent of genome complexity [3, 8].

The objective of this research is to investigate alternative strategies for mapping complex traits. The advent of whole genome expression profiling technology has enabled investigators to combine functional genomics with genetic variation such as SNPs to inform the choice of possible candidate genes for further study. In addition, the recognition of the scope and magnitude of common structural variation within the human genome has revealed the possibility that genomic variation other than SNPs could impact complex trait susceptibility. Although the role of altered DNA structure in rare Mendelian diseases and genomic disorders has long been established, the dynamic range of non-disease variation in the human genome has only recently been appreciated. Studies mapping large scale DNA polymorphisms across the genome [8-15], have shown that much of the common variability in DNA structure normally impacts as much as 25% of the genome and is located in many known disease regions. The advent of high throughput technologies allows current gene mapping studies to integrate the results from multiple strategies to identify genes involved in complex traits (Figure 1). In this study, genes involved in age-related cognitive decline as defined by Alzheimer's disease (AD) were identified through the use of current high-throughput technologies. Gene profiling was used to identify functional candidates and associated pathways implicated in neuropathological phenotypes related to cognitive decline. Whole genome association (WGA) techniques were investigated to assess both the low level characteristics of the genotyping algorithms and the impact of DNA structural variation on the previous gene expression results.

B. Gene expression profiling to identify functional candidate genes

Finding the genes involved in a complex phenotype such as healthy brain aging is challenging due to the biological complexity of the underlying genetic and environmental components. A primary challenge is presented by the heterogeneity of the phenotype itself. Individuals exhibit broad variation in the ability to maintain cognitive function during the aging process. Clinically significant cognitive decline in the elderly is most commonly caused by Alzheimer’s disease (AD). Diagnostic neuropathological features of AD include extracellular amyloid plaques and intracellular neurofibrillary tangles (NFTs). However, there is considerable neuropathological heterogeneity across individuals with clinical AD and individuals with no clinical signs of dementia, making division into “cases” and “controls” based on neuropathology problematic. In particular, there is tremendous variability in the relationship between the amount and location of AD neuropathology in the brain and the clinical manifestation of AD symptoms [16, 17]. Individuals with robust cognitive function may tolerate high levels of brain tissue injury presumptively indexed by amyloid plaques and NFTs, while others demonstrate loss of cognition with similar or even lower levels of lesion burden.

According to cognitive reserve theory, individuals differ in their capacity to maintain normative cognitive function and, accordingly, those with greater capacity are better equipped to delay or circumvent the damaging effects of brain lesions that in other less equipped individuals, lead to clinical manifestations of AD. The theory postulates that this natural variability across individuals is due to differences in neural processing mechanisms [17]. The physiological basis of this mechanism is unknown, although it is likely to reflect environmental as well as genetic factors [18, 19]. Genetic variations can contribute to individual differences in normal cognitive function. Interaction between these genetic differences and environmental factors over the lifespan can amplify variation in cognitive function later in life.

There is growing evidence that variation in the quantity of a gene product, rather than simply presence or absence of product, can be responsible for the subtle effects of complex traits [20-22]. Several recent studies have shown that variation in gene expression is heritable [23-25] and can be mapped as a quantitative trait [25].

We performed whole genome expression profiling of RNA obtained from frontal cortex of clinically non-demented and AD subjects to identify genes associated with brain aging and cognitive decline. Genetic mapping information and biological function annotation were incorporated to highlight genes of particular interest. The candidate genes identified in this study were compared with those from two other studies in different tissues to identify common underlying transcriptional profiles. In addition to confirming sweeping transcriptomal differences documented in previous studies of cognitive decline, we present new evidence for up-regulation of actin-related processes and down-regulation of translation, RNA processing and localization, and vesicle-mediated transport in individuals with cognitive decline.

C. Low –level analysis of high-throughput genotyping

Studies investigating the relationship of specific DNA variants to human complex traits rely both on large numbers of subjects and large numbers of genetic markers located across the entire genome. Because of these requirements, the technology necessary to economically identify and characterize genetic changes contributing to complex traits has been available only recently[26, 27]. Arrays consisting of SNPs at high density across the genome have been used successfully to detect genetic variation involved in complex human disease and drug susceptibility [3, 28-30]. Paired with the escalating technology to manufacture arrays with increasingly dense SNPs interrogated on one array is a corresponding rise in the challenge to distinguish genotypes for all SNPs across the array in all samples. Genotyping algorithms for hybridization based techniques are designed to convert the raw signal intensity values obtained from the array into SNP genotypes usable for analysis. Inaccurate and incomplete genotype calls introduce variability into a data set resulting in increased rates of discordance and loss of power to answer higher level biological questions. Therefore, the effectiveness of genotyping algorithms shapes the interpretation of experiments.

Affymetrix genotyping arrays consist of multiple probe sequences for every SNP interrogated [27] and algorithms vary in the statistical methodology used to summarize probe level data. Similar to developments in gene expression microarray analysis [31-33], assessment of the low-level performance of high throughput genotyping algorithms will allow possible improvements to be identified and help reduce the loss of data across samples. With this in mind, we investigated the factors that might affect the performance of two genotyping algorithms for the commonly used Affymetrix GeneChip 100K Mapping Array genotyping platform.

The Affymetrix GeneChip 100K Mapping Array Set consists of two chips designed to genotype a total of 120,000 SNPs. Each chip is hybridized with a DNA sample digested by either the XbaI or HindIII restriction enzyme. There are 40 probes specific to each SNP target. A probe is a 25 bp oligomer centered at the SNP site (Figure 2). Probes are organized into probe sets consisting of a quartet that interrogates each allele (A and B) on each strand with both a perfect match 25-mer complementary to the SNP sequence and a mismatch with the center 13th base substituted. An additional four probe sets are similar in configuration but are center shifted from the SNP such that the center 13th base is either 1, 2, 3 or 4 base pairs away from the SNP site. There are five quartets tiled each for both the sense and antisense strands for a total of 10 probe quartets specific to each SNP.

The Affymetrix GeneChip 100K Mapping Array Set Dynamic Modeling (DM) algorithm is implemented in the Affymetrix GType software [34]. The DM algorithm processes each chip independently and each SNP one at a time. Log likelihood estimates of the four possible genotype models (AA, AB, BB or NoCall) are calculated for each probe quartet (Figure 3). The ten quartets are combined and evaluated for significance using the Wilcoxon signed rank test. The model with the lowest p-value is the genotype call for that SNP.

The multi-chip genotyping algorithm (BRLMM) recently introduced for the Affymetrix GeneChip 500K Mapping array set (http://www.affymetrix.com/support/technical/product_updates/brlmm_algorithm.affx) is a Bayesian modification of the RLMM algorithm [35] whereby information from all SNPs across all chips in an experiment are used to model spatial clusters in a classification approach to genotype calling. Non-biological variance is reduced through normalization across both chips and probe sets.

The effectiveness of each algorithm in accurately calling SNP genotypes was originally determined by comparisons with HapMap genotyping data (http://www.affymetrix.com/support/technical/product_updates/brlmm_algorithm.affx). However, the influences of chip specific and experimental specific characteristics on algorithm performance were not investigated. In this study, we used a combination of simulations and sample genomic DNA from control and Alzheimer's disease patients to investigate the sensitivity of these genotyping algorithms to chip type, background noise, experimental variability and sample characteristics.

D. Integration of genome wide copy number and gene expression

Gene expression levels vary in individuals within and across populations [24, 25, 36-38]. Identification of specific genetic variants contributing to variation in gene expression has typically been focused on SNPs. Mapping studies of variation in gene expression levels as quantitative traits have been focused on the impact of SNPs and have led to the identification of cis (genetic variants located on the same homolog as the transcript) and trans (genetic variants located elsewhere) acting SNPs. The recent discovery of the extent of copy number variants (CNVs) across the human genome indicates that normal genetic variation encompasses a wider range of genomic architecture than previously thought and raises the question of whether these variants could influence complex traits [39-43].

Differences in copy number can impact transcription in several ways [44]. Amplification or deletion of large stretches of DNA encompassing multiple genes may affect expression levels of all genes in the region. Smaller copy number variable regions could impact dosage sensitive genes or unmask a recessive allele on the homologous chromosome. CNV that overlap dosage sensitive genes can disrupt the genes leading to various outcomes such as reduced expression or altered transcripts. Deletion of regulatory elements in CNVs near genes could reduce gene levels or unmask regulatory mutations. A combination of several CNVs or CNVs interacting with specific SNPs could lead to altered expression whereas each one alone does not.

Differences in copy number are known to affect the expression of single genes [45-51] as well as those within genomic regions altered in tumors [52]. However, gene expression levels do not automatically reflect copy numbers due to the complex regulatory mechanisms invovled in gene expression[53-55]. It is unclear to what extent changes in copy number affect normal gene expression [56].

We set out to investigate the impact of copy number differences on the results of differential gene expression experiments. Our previous study identified differentially expressed genes in subjects with cognitive decline relative to non-demented subjects [57]. We used DNA from these subjects to identify copy number variants and compared three copy number estimation programs to determine the effectiveness of each algorithm in identifying CNVs in our samples.

Copy number variation in non-tumor tissues tends to be smaller in amplitude and more focused, therefore encompassing fewer SNPs. Current algorithms for copy number estimation are optimized for the types of copy number changes seen in cancer tissues [58]. It is unknown to what extent these algorithms will detect the type of copy number variability generally seen in normal individuals. To determine the contribution of CNVs to gene expression in our study, we compared the regions of copy number variability to the differentially expressed genes found in the previous study.

Gene mapping in cognitive decline was investigated through the use of high-throughput methods. Chapter II describes the results of a whole genome gene profiling experiment in subjects diagnosed with AD and controls which addresses the phenotypic heterogeneity seen in this disease by stratifying AD subjects by severity of neuropathology. Secondly, a whole genome sampling assay (WGSA) using the same samples as in the gene expression profiling study was performed. The purpose of the whole genome genotyping was to determine the effect that copy number variation might have on differential gene expression. Chapter III details the comparison and considerations of algorithms used to summarize the probes and obtain genotypes on the Affymetrix 100K Gene Mapping Set. Chapter IV details the comparison and considerations of the algorithms for determining copy number differences between samples and the impact of copy number differences on gene expression. The appendix describes a method for integrating diverse biological information into a single list of genes prioritized by a weighting scheme.

Chapter I. Figures and Table

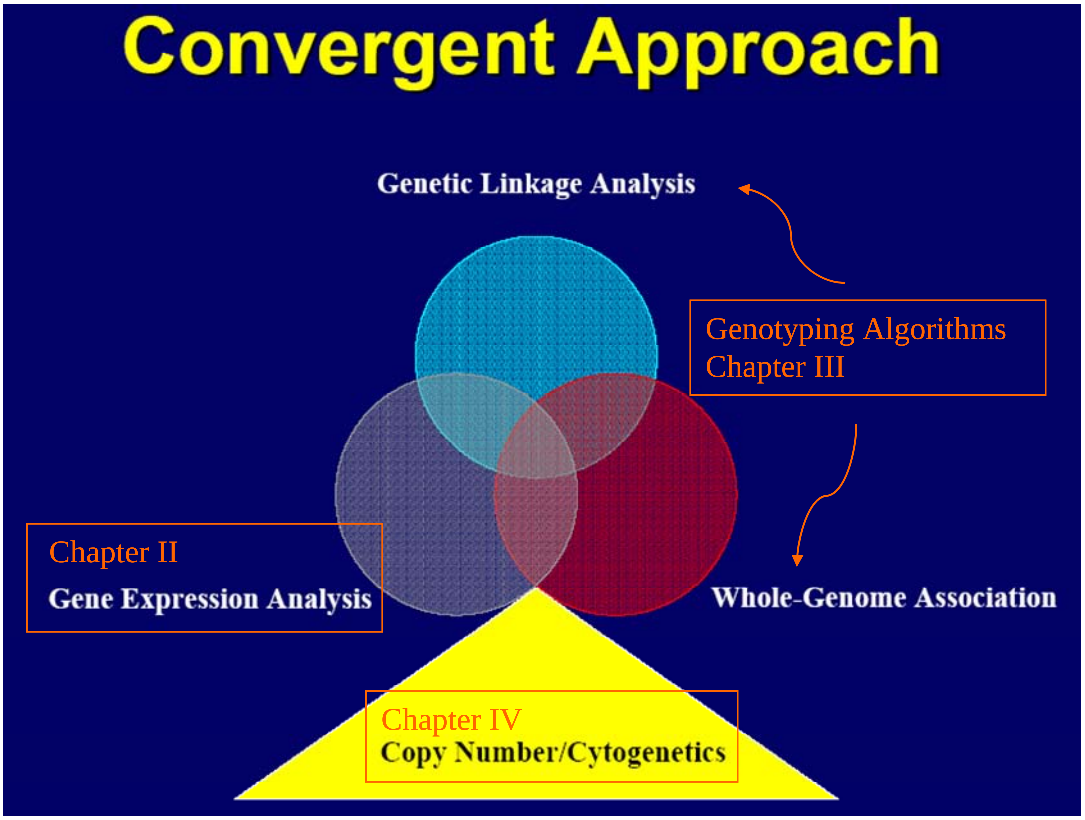


Figure 1. Integration of strategies for gene mapping in complex traits.

(modified from Frank A Middleton)

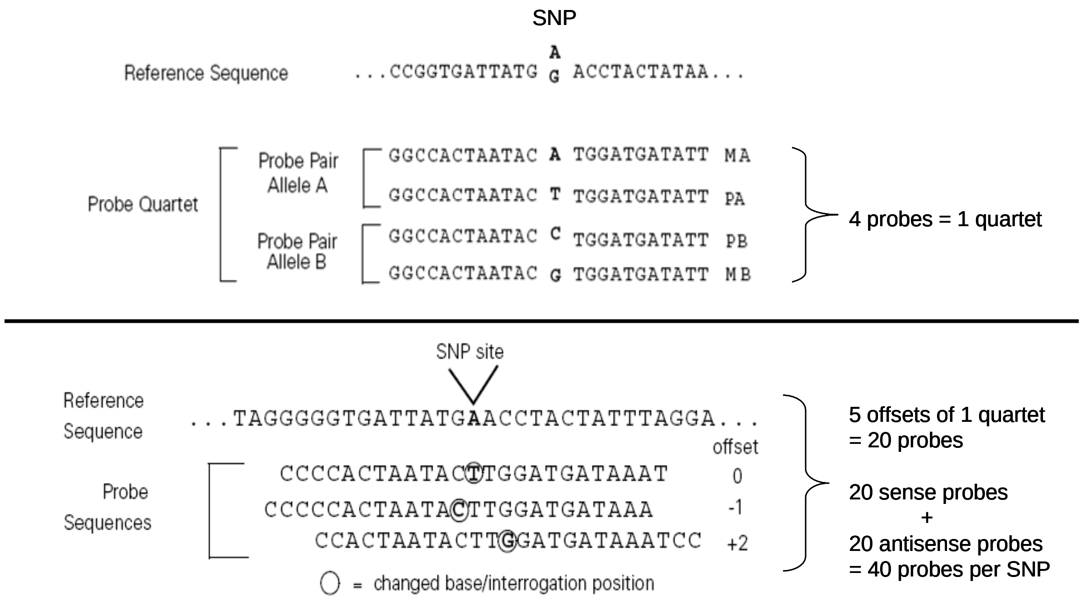

SNP
Reference Sequence
...CCGGTGATTATG A G ACCTACTATAA...
Probe Quartet
Probe Pair Allele A
GGCCACTAATAC A TGGATGATATT MA
GGCCACTAATAC T TGGATGATATT PA
Probe Pair Allele B
GGCCACTAATAC C TGGATGATATT PB
GGCCACTAATAC G TGGATGATATT MB
4 probes = 1 quartet
SNP site
Reference Sequence
...TAGGGGGTGATTATGAACCTACTATTTAGGA...
offset
Probe Sequences
CCCCACTAATACTTGGATGATAAAT 0
CCCCCACTAATACTTGGATGATAAA -1
CCACTAATACTTGGATGATAAATCC +2
○ = changed base/interrogation position
5 offsets of 1 quartet = 20 probes
20 sense probes + 20 antisense probes = 40 probes per SNP

Figure 2. Probe set design for the Affymetrix 100K mapping GeneChip. Genotypes at each SNP are a summarization of 40 probes. The perfect match (PM) and mismatch (M) probes for the two alleles A and B (4 probes/1 quartet). There are 5 quartets per SNP which have the center base offset from the original SNP site (5 quartets = 20 probes). Both the sense and antisense strand are interrogated (20 probes x 2 strands = 40 probes). (Modified from GType Mannual, Affymetrix, Inc.)

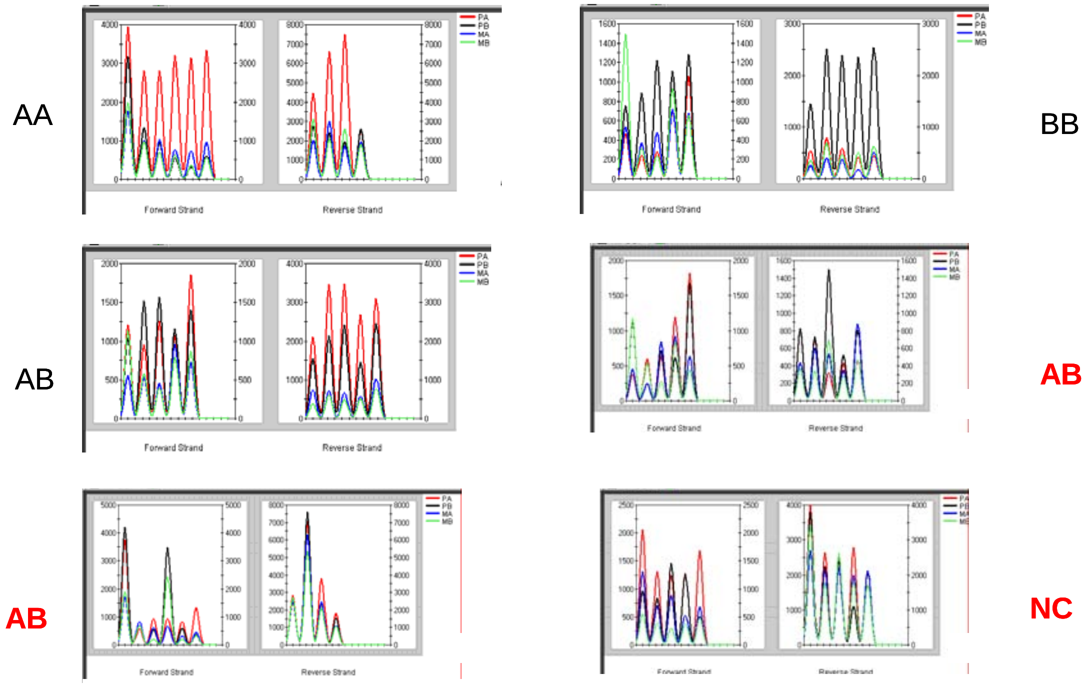


Figure 3. Individual probe intensities for six different probe sets on the Affymetrix 100K GeneChip. Genotype calls are that of the DM and BRLMM algorithms. Red lines are PM A, black lines are PM B, blue lines are MM A, and green lines are MM B. Genotypes in red are BRLMM genotypes and were called NC (not called) by the DM algorithm.

Chapter II. Translational gene mapping of cognitive decline.

Introduction

The ability to maintain cognitive function during aging is a complex process subject to genetic and environmental influences. Alzheimer's disease (AD) is the most common disorder causing cognitive decline among the elderly [59]. Among those with AD, there is broad variation in the relationship between AD neuropathology and clinical manifestations of dementia [17, 60]. These differences in protection from the effects of AD neuropathology may be due to genetic differences at several levels including the expression of gene products. We suggest that differences in expression of genes in neural processing pathways are responsible for differences in the maintenance of cognitive function, and at least in part account for an important component of cognitive reserve.

To address this hypothesis, we performed whole genome expression profiling on a set of well-characterized, clinically non-demented and AD subjects in order to identify genes, or gene pathways, that contribute to cognitive decline. Subjects were stratified into four groups based on cognitive status prior to death (non-demented or AD) and neuropathological status defined by three categories of intracellular neurofibrillary tangle (NFT) burden which is determined by the quantity of NFT in specific areas of the brain (Braak stage I/II, III/IV, and V/VI [61]) (Figure 1A). Non-demented subjects were represented in all three Braak-stage categories, whereas AD subjects were represented only in Braak stage V/VI. We designed three comparisons to test three hypotheses (Figure 1B). In the first comparison, we postulated that all non-demented subjects, taken as a whole (Groups 1, 2 and 3), would exhibit different gene expression profiles compared to AD subjects (Group 4), irrespective of NFT burden. We refer to this as the

Extreme Cognitive Phenotypes Hypothesis (Hypothesis I). In the second comparison, we proposed that individuals with lower NFT burden (Braak stage I/II and III/IV, Groups 1 and 2) would display different expression profiles than those with higher NFT burden (Braak stage V/VI, Groups 3 and 4), irrespective of cognitive ability. We refer to this as the Neuropathologic Process Hypothesis (Hypothesis II). In the third comparison, we postulated that expression profiles in non-demented subjects with a high NFT burden (Group 3) would differ from those in AD subjects with similar NFT pathology (Group 4). We refer to this as the Cognitive Reserve Hypothesis (Hypothesis III).

We interpret our gene expression results in the context of prior evidence from genetic linkage studies and biological function annotations to identify possible candidate susceptibility genes. Furthermore, since genes that are differentially expressed across tissues involved in AD pathology would provide valuable insight into common underlying genetic mechanisms in brain aging, we compared genes identified in this study, using frontal cortex, with genes identified in two other expression studies using hippocampus [62, 63] and entorhinal cortex [62, 63]. Genes that were differentially expressed across the three studies, emphasizing common themes of pathology underlying dementia, are key candidates for further studies of genetic risk factors for cognitive decline.

2. Materials and methods

2.1 Patient and control samples

Postmortem human brain tissue comprised primarily of gray matter from frontal cortex was obtained from the neuropathology core of the NIA-Layton Aging and Alzheimer's disease Center, Oregon Health & Sciences University (OHSU). All subjects were characterized based on specific clinical and neuropathologic criteria [64] through studies performed by the NIA-Layton Aging and Alzheimer 's disease Center. An extensive collection of clinical data, including cognitive and functional measures, and neuropathologic data was available for all subjects. Testing included annual cognitive, functional and neuropsychological examinations. At autopsy, portions of the brain were frozen at –80^{o}C and the remainder was prepared for histological examination by fixation in 10% formalin. All subjects were scored for neuritic amyloid plaques and neurofibrillary tangles according to NIA-Reagan criteria [64, 65].

All subjects met the following minimal criteria for study inclusion: post-mortem interval < 24 hours, neurological examination within one year of death, Caucasian, non-detectable cancer metastases, and minimal degradation of brain-derived RNA for microarray analysis (see below). AD subjects were also required to have age at onset > 70 years and a clinical diagnosis of Probable AD. AD subjects with a coexisting neuropathologic diagnosis of Parkinson's disease, Lewy Body Dementia or Frontotemporal Dementia were excluded from the study. Non-AD subjects were required to have a clinical diagnosis of "non-demented", a CDR score of 0 and a Mini-Mental State Examination score (MMSE) > 25 (Table 1). Braak stage [61] was used to further define all subjects with respect to severity of neurofibrillary tangle burden (Figure 4A). The study sample comprised fourteen subjects (7 male, 7 female). Average age at death was similar across all groups (89.7 - 93.6 years). Non-demented subjects had an average MMSE score of 28.4; AD subjects had an average MMSE score of 14.4 (Table 1).

2.2 RNA isolation and hybridization

Approximately 500 mg of fresh frozen brain tissue from each individual was processed for total RNA using the RNeasy kit (Qiagen Inc., Valencia, CA). RNA quality was assessed by UV absorbance measurement and electrophoresis on RNA NanoChips using the 2100 Bioanalyzer (Agilent, Palo Alto, CA). Samples were considered acceptable for labeling and further processing if UV260/280 ratios were greater than 1.7 and Bioanalyzer profiles showed minimal degradation. For determination of degradation status, Bioanalyzer profiles were referenced to a simultaneously processed control of high quality RNA whose profile correlated with good performance on an Affymetrix GeneChip array (Gene Microarray Shared Resource, OHSU).

Two ug of total RNA from each subject was amplified and labeled using the AMC one cycle cDNA, Affy IVT amplification/labeling protocol following manufacturer’s instructions (Affymetrix Inc., Santa Clara, CA). Labeled targets were hybridized with Affymetrix GeneChip HG-U133 Plus 2.0 arrays. These arrays contain 47,000 transcripts spanning the entire human transcriptome. Sample labeling and array hybridizations and processing were performed in the Affymetrix Microarray Core, Gene Microarray Shared Resource, OHSU.

2.3 Realtime RT-PCR

Confirmation of array results was performed using TaqMan chemistry in qRT-PCR. Phenotypic heterogeneity as well as RNA quality profoundly effect gene expression levels. Two additional non-demented subjects, conforming to the same rigorous phenotypic criteria, were included with the original set of subjects, in order to substitute for two non-demented subjects for which RNA had degraded in the interim between the microarray analysis and the validation procedure. Polyadenylated mRNA from the total RNA isolated from frontal cortex was reverse transcribed (Transcriptor RT, Roche Diagnostics Corp, IN) using oligo dT primers (Invitrogen, CA). Specific primers corresponding to the short form of ITSN1 (Hs00495035_g1, Applied Biosystems, TX) were combined with cDNA and dNTPs in a master mix (FastStart DNA Master Hybrid Probes, Roche Diagnostics Corp, IN) and amplified by PCR in a SmartCycler (Cepheid, CA). Human mRNA (Ambion, Inc, Tx), treated in the same manner was used as the control sample. Because standard housekeeping genes displayed variable expression levels across sample groups, qRT-PCR reference genes were chosen from the results of the HG-U133 Plus 2.0 arrays. Two different genes (POL2RF, RTN2) were chosen based on their lack of differential expression across groups and for their relative levels of expression similar to ITSN1 in the non-demented group. Samples were run in triplicate and the efficiency for each reaction was determined based on linear regression analysis of the exponential phase of the reaction [66] Relative gene expression of ITSN1 to each reference gene was calculated using the efficiencies and crossing threshold (Ct) of each reaction [67]:

$$\text{Relative Ratio} = \text{Efficiency}_{\text{ITSN1}}^{(\text{Ct control} - \text{Ct sample})} / \text{Efficiency}_{\text{reference}}^{(\text{Ct control} - \text{Ct sample})}$$

2.4 Statistical analysis

An overview of the entire analytical work flow is provided in Figure 2. Statistical analyses were performed in the R v2.0.1 system for statistical computation ([68], http://www.R-project.org). Packages included in the Bioconductor v1.6 suite of analysis tools for genomic data [69] were utilized for specific analyses, as well as custom scripts.

Hybridized arrays were rigorously evaluated for quality using the *Affy* package v1.5.8 [70] of the Bioconductor project. Computer-generated graphs of the hybridization intensities across the chips allowed a visual assessment of the consistency of the hybridization reaction. Model-based normalization procedures were used to correct for systematic biases. Scatter plots [71] were used to compare the shapes of the distributions before and after normalization. Post-normalization residual plots were used to assess the model fit across all arrays.

Systematic errors cause technical variation which reduces the power of an array experiment to elucidate true biological variation. To minimize the impact of this variation on data analysis and biological interpretations [31], we used two different low level analysis approaches. Each data set was analyzed separately, allowing us to compare the impact of the low level routines on the downstream analysis. The Robust Multi-chip Analysis (RMA) [31] is a model-based pre-processing algorithm used to correct for probe-level differences. RMA in the *Affy* package was performed on log-transformed hybridization intensities using RMA background correction, quantile normalization and median polish as a summary statistic.

The Variance Stabilization and Calibration (VSN) [33] algorithm of the *Affy* package is a model-based normalization algorithm that specifically transforms the data such that the variance is independent of the mean intensity. The VSN algorithm was performed on intensity values and summarized using the median polish algorithm.

RMA and VSN processed data sets were analyzed to identify putative differentially expressed genes using Analysis of Variance (ANOVA) with the Linear Models for Microarray data analysis package (LIMMA v1.8.10) [72] of the Bioconductor project. Individual linear models were fitted for each transcript across the groups. The first two hypotheses were formally tested as planned comparisons within this framework.

Because each transcript is tested separately, and given the large number of transcripts on the array, the false positive error rate increases dramatically. Therefore, the q-value statistic [73], a minimum measure of the False Discovery Rate (FDR), was used to correct for multiple testing. The FDR is the number of predicted false positive results out of all significant tests. This measures the significance of each gene, taking into account that thousands of genes are being tested. Q-values were calculated from p-values generated in the LIMMA analysis using the QVALUE package v1.1 for R [73].

For each analysis, the final list of putative differentially expressed genes was defined as those probe sets with a q-value < 0.10 that occurred in both the RMA and VSN normalized data sets, in order to balance statistical rigor with maximal identification of candidate genes and given the discovery framework of this study.

It is noted that in the original experimental design, all three hypotheses were to be formally tested. However, during the QA/QC process for sample quality and hybridization, the loss of samples resulted in the third contrast being underpowered, leading to a different statistical approach for this comparison. Vector Projection is a dimension reduction technique for the rapid identification of genes with particular patterns of expression across groups (Terry Speed, Department of Statistics, University of California, Berkley, and Genetics and Bioinformatics, Walter and Eliza Hall Institute Australia; and Ingrid Lonnstedt, Department of Mathematics, Uppsala University, personal communication to S. McWeeney, [74]). It is useful as an initial exploratory data analysis tool, particularly when limited sample sizes preclude formal trend analysis, as was the case with Hypothesis III. Each gene has a vector of its normalized expression values across time. These values are projected onto the space spanned by the pattern of interest (vector of coefficients or weightings for group). In this case, the pattern of interest was a contrasting expression pattern between Cognitive Reserve (CR) and the other groups (i.e., identify genes up-regulated in CR and down-regulated in the other groups, or vice versa). Projection scores in the extreme tails of the normal Quantile-Quantile (QQ-plot) were used to identify transcripts with the best fit to the pattern of interest. The significance level was set at 0.1 for q-values in all expression analyses due to the gene discovery framework of this study.

2.5 Determination of biological significance

All transcripts on the array were annotated for gene name, function, and chromosome location using NetAffx (http://www.affymetrix.com/analysis/index.affx, NCBI build 35). These annotations were then used for subsequent downstream analysis. For overrepresentation analyses (linkage, chromosome bands, GO), all significance levels were set at 0.05.

Transcripts that were differentially expressed in non-demented versus AD subjects (Hypothesis I) were analyzed for overrepresentation in specific chromosome regions in two ways. First, transcripts were annotated for cytogenetic bands and a χ^2 test of independence was performed to determine if there was evidence for association of transcript expression and cytogenetic band location. Secondly, transcripts were examined for their presence in a chromosomal region known to be linked or associated with AD from previous studies. Concordant linkage/association regions were identified [60]. The number of differentially expressed transcripts located in these regions was compared to the number of transcripts in these regions on the Affymetrix HGU133 Plus 2 GeneChip array, using a one-tailed Fisher's exact test to determine if the number of differentially expressed transcripts located in each region was greater than that expected by chance.

Differentially expressed genes identified by ANOVA (q-value < 0.1) were assigned to Biological Process categories of the Gene Ontology (GO) Consortium (http://www.geneontology.org/ August, 2005). The GO is an international effort to define genes and their products using a controlled vocabulary. We used GOSTAT [75] to assess representation of differentially expressed genes in GO Biological Process categories. Identification of pertinent pathways depends on the availability of annotations mapped to the probe set. Differentially expressed genes were compared to all genes on the HG-U133Plus2 GeneChip array, using a 2 x 2 contingency table and counting the number of appearances of each category for differentially expressed genes versus reference genes. The probability that differentially expressed genes fall within a category more often than what would be expected by chance was calculated by χ^2 (Fisher's Exact test if the counts within a category are below 5). FDR was used to correct for multiple testing by controlling for interdependencies among the categories [76] given the hierarchical nature of the GO Ontology.

The Kyoto Encyclopedia of Genes and Genomes (KEGG) database (http://www.genome.jp/kegg/, September, 2005) was used to classify differentially regulated genes into canonical pathways for biological interpretation. Transcripts were annotated for their presence in a KEGG pathway and the significance of the number of genes differentially expressed in each pathway was determined by a one-tailed Fisher's Exact Test.

2.6 Identification of genes in common across tissues

Comparison of our study with two previously published gene profiling experiments [62, 63] was used to identify genes that would reveal common pathophysiological mechanisms. First, differential gene expression related to cognitive decline was determined by combining transcripts differentially regulated in the comparisons of AD versus non-demented subjects in hippocampus Blalock, 2004 #74, entorhinal cortex [63] and frontal cortex (this study). Because Blalock, et al Blalock, 2004 #74 used the Affymetrix HG_U133A GeneChip array, we used a subset of data from the other two studies that corresponded to the Probe IDs found on the Affymetrix HG_U133A GeneChip. Significance was set at $p < 0.1$ for each data set and the intersection of Affymetrix Probe Ids was defined as the set of transcripts in common. It is noted that we cite p-values rather than q-values for this component as that is what was reported by the other studies. Because there are often multiple transcripts mapping to one gene on the GeneChip array, we also generated a data set of the intersection of differentially expressed genes in common among the three experiments using the annotated gene symbol. Secondly, we compared the transcripts involved in NFT formation (Hypothesis II) with the differentially expressed transcripts obtained by Dunckley et al. [63] from neurons without NFTs from AD subjects versus adjacent neurons with NFTs. The final data set of transcripts involved in NFT formation was defined to be the intersection of Affymetrix HG_U133Plus2 Probe IDs differentially expressed in both data sets.

3. Results

Putative differentially expressed transcripts were identified based on Hypothesis I (extreme cognitive phenotypes) and Hypothesis II (NFT formation) in order to identify genes involved in different, but overlapping, features of age-related pathological processes. Particular attention was focused on differentially expressed genes in chromosomal regions shown to be linked or associated with AD in previous studies [74]. In addition, we used the Gene Ontology (GO) Biological Process categories to identify cellular events influenced by the differentially expressed genes associated with cognitive decline. For the probe sets identified in our analyses, only a subset had available GO annotations (Table 2). The overall level of available gene annotation was 32% of the unique genes annotated for GO Biological Process terms. Subsequent analyses are dependent on these annotations.

3.1 Extreme cognitive phenotypes

We identified 8346 transcripts, representing 5096 genes, that were differentially expressed ($q < 0.1$) between non-demented and AD subjects (Hypothesis 1, Figure 4B) (Supplemental Table 1). Cytogenetic band annotations were available for 6857 transcripts, of which 339 (4.9%) were located in regions with higher numbers of differentially expressed transcripts than expected by chance ($p < 0.05$, Supplemental Table 2). Ten cytogenetic bands contained more differentially regulated transcripts than would be expected by chance ($p < 0.05$, Supplemental Table 2).

All 8346 differentially expressed transcripts were annotated for location in a genomic region shown previously to be linked or associated with AD [60] (Supplemental Table 1). Of the total 8346, 873 transcripts were located within the sixteen linkage regions (Table 3). Of the transcripts up-regulated in AD, 264 are located in linkage regions. The most significant up-regulated transcript (35776_at) is the short form of Intersectin 1 (ITSN1), located in linkage region 21q22.1-q22.2. Affymetrix probe sets allow comparison of specific alternative transcripts in differential gene expression. For this gene, a different probe set interrogating the short form is also significantly up-regulated (209297_at, q = 0.01) but the long form is not differentially expressed (q = 0.25). These results were confirmed by qRT-PCR (Figure 3) where the short form of ITSN1 is up-regulated in AD relative to two different reference genes (p = 0.009 and p= 0.025). The long form was not differentially expressed (data not shown). The most significant down-regulated transcript, ATP6V1G2, is also located in a linkage region, 6p21.3.

Cognitive decline, represented by AD subjects in our analysis, reveals a massive restructuring of cellular physiology (Table 4). Many of the most significant up-regulated categories are related to regulation of cellular functions. Categories related to transcription and its regulation, including chromatin modification, are among the most highly represented. Transcripts for actin-related processes and phosphate transport are also up-regulated.

Widespread down-regulation occurs in energy pathways and nucleic acid-related categories. Additionally, secretory pathways, RNA-related categories including splicing and mRNA processing, many pathways related to protein metabolism including folding, localization, targeting, transport and translation are down-regulated. Transcripts from genes involved in mitochondrial physiology are also down-regulated.

We utilized the KEGG database to place the differentially regulated genes into canonical pathways (Table 5). Of the 5096 differentially expressed genes, 226 were found to be clustered at levels greater than what would be expected by chance in 14 KEGG pathways. Of these, nearly half (45.5%) are involved in energy metabolism (oxidative phosphorylation, ATP synthesis, carbon fixation and CO_2 fixation). An additional 18% are involved in genetic information processing (transcription, translation and protein degradation). Carbohydrate (12.8%), amino acid (9.7%) and lipid (2.2%) metabolism are also represented. The percent of differentially expressed genes in each pathway (% abundance) varies from 37.2% - 71.4%. The pathway with the greatest percentage of differentially expressed genes (synthesis and degradation of ketone bodies) has the lowest number of total genes in the pathway. The pathway containing the greatest number of differentially expressed transcripts (50.7%) was oxidative phosphorylation.

3.2 Neurofibrillary tangle formation

We identified 528 transcripts, representing 492 genes, which were differentially expressed ($q < 0.1$) between subjects with low NFT pathology and those with high NFT levels (Hypothesis II, Figure 4B) (Supplemental Table 3). Of these, 98.9% were also differentially regulated in the Extreme Cognitive Phenotypes comparison. The six genes unique to Hypothesis II are close to the 0.1 threshold for significance (data not shown). A total of 49 transcripts were located in linked regions (Supplemental Table 3).

Overrepresentation in GO Biological Process categories reflected the dependence on current annotation. Specifically, the significant categories were dominated by a small number of well studied genes with pleiotropic effects (data not shown).

3.3 Cognitive reserve

Vector projection analysis allowed initial determination of putative candidate genes involved in cognitive reserve. Eleven transcripts, all located outside known AD linkage regions, were identified as possible candidates (Table 6). Of these, only one (GSTT1, involved in glutathione metabolism) was also differentially regulated in the Extreme Cognitive Phenotypes comparison. All other genes are unique to the Cognitive Reserve analysis.

3.4 Identification of common themes related to cognitive decline

We compared genes identified in this study, using frontal cortex, with genes identified in two other expression studies using hippocampus and entorhinal cortex. Blalock et al [62] compared hippocampal gene expression in non-demented and AD subjects stratified by severity of disease as measured by NFT count and MMSE scores. Dunckley et l [63] used laser capture microdissection (LCM) to obtain RNA from neurons in entorhinal cortex, and then compared gene expression patterns in NFT-containing neurons and adjacent NFT-free neurons in AD subjects. Neurons without NFTs were also obtained from non-demented subjects for comparison.

In order to identify genes common to the underlying process of cognitive decline, we combined the data sets across the three different tissues (Table 7A and Table 8). Pairwise comparisons for all transcripts on the HG_U133A GeneChip array showed similar concordance with our data and either of the other data sets. Concordance rates among any two data sets varied between 7.1% and 20.8%. A total of 174 transcripts were concordant (FDR 10%) across all three data sets. More stringent criteria (FDR 5%) resulted in a loss of 30% of those transcripts. The overall concordance rate for differentially-regulated transcripts across all three data sets ranged from 1.0% - 3.9%. Of the 18 transcripts located in linkage regions (Table 8), six are involved in intracellular transport (ITSN1, ATP6V1G2, SYNJ1, SYNCRIP, DIRAS2) and three are related to mitochondria (ATP5J, ATP5C1, MRPS10). GO category analysis of the entire concordant transcript IDs demonstrated that the most significantly overrepresented GO category for up-regulated genes was signal transduction (data not shown). Down-regulated transcripts were most notably overrepresented in energy pathways and carbohydrate metabolism (data not shown). If the differentially expressed genes are mapped to gene symbol ID, the number of genes common to all three data sets increases (8.1-36.6%, FDR 10%) (Table 7B and Supplemental Table 4).

Dunckley et al [63] compared neurons with and without NFTs in AD subjects in order to investigate NFT formation. We compared low Braak stage subjects with high Braak stage subjects regardless of cognitive function for the same purpose. Transcripts differentially expressed in both data sets showed 39 (9.8%) concordant transcripts (Supplemental Table 5). Most are down-regulated in subjects with higher numbers of tangles (74.3%).

4. Discussion

Results of our human transcriptome profiling confirm many of the sweeping transcriptional differences associated with cognitive decline that have been previously documented, and implicate genes involved in transcriptional regulation, energy pathways, ion homeostasis dysregulation, apoptosis, and synaptic activity [62, 77-79]. In addition, our results reveal significant up-regulation of actin-related processes and down-regulation of translation, RNA processing and localization, and vesicle mediated transport (Tables 4 and 5). This study identifies candidate genes, located in linkage regions, which had not been previously implicated in cognitive decline.

One difficulty with microarray results is that, because biochemical networks connect multiple physiological processes, a plausible biological mechanism for the implication of many genes can often be suggested. This is compounded when studying a complex trait impacting multiple cellular functions. We found that interpreting gene expression results in the context of genetic mapping studies and functional annotation allowed a more informed approach to identifying candidate genes in brain aging.

4.1 Extreme cognitive differences

We localized differentially expressed genes in healthy aging versus cognitive decline with reference to cytogenetic band annotations. In 14 genomic regions, more transcripts were differentially expressed than would be expected by chance (Supplemental Table 2), indicating possible co-regulation of genes in these regions by trans-acting factors. We identified functional changes of genes located in known AD linkage regions through differences in expression to identify cis-acting DNA polymorphisms. AD linkage regions did not overlap with the 14 genomic regions, indicating that the greater number of genes located within linkage regions was not coordinately regulated by trans-acting factors. The majority (87%) of transcripts were not found in linkage regions. However, differentially regulated genes located within the known AD linkage regions may contain cis-acting DNA polymorphisms that affect their gene expression and contribute to the linkage signal. Our results identified 873 possible candidate transcripts.

Biological annotation of these transcripts revealed that a number of these genes are involved in synaptic dysfunction, which has been shown to be an early process in cognitive decline. Synapse loss correlates positively with cognitive decline and indeed may occur prior to clinical signs [80]. Enlarged endosomes appear early in the course of AD pathology and are not present in healthy aging [81]. While many synapse-specific genes and vesicle-mediated transport genes are generally down-regulated in our study, we have identified a significantly up-regulated transcript, ITSN1, which is located in linkage region 21q22 (Supplemental Table 1). ITSN1 has not been studied in cognitive decline, although it has been postulated that ITSN1 might affect APP processing [82] and vesicular trafficking in AD [83]. Analysis of the other published data sets also identified ITSN1 as consistently up-regulated (Supplemental Table 4).

ITSN1 is a scaffold protein involved in synaptic vesicle recycling [84] and caveolae internalization [85]. Overexpression of ITSN1 blocks clathrin-mediated endocytosis [86], internalization of caveolae [85] and Ras activation [87] (Figure 4). Inhibition of endocytosis has been shown to increase soluble APP alpha release [88-90]. The fundamental significance of ITSN1 is its role in linking the endocytic machinery at the synapse with both the actin cytoskeleton and signal transduction pathways. Signaling pathways are regulated through ITSN1 binding of SOS and activation of RAS [87] and Elk1 activation through a RAS-independent process involving JNK[91]. Rho/Ras signaling is related to actin cytoskeleton through the protein kinase ROCK1 [92] that is also up-regulated in AD brain tissue (Supplemental Tables 1 and 4). The consistent findings across expression studies and the functional consequences of its overexpression provide compelling evidence for a central role for ITSN1 in the pathogenic mechanisms of cognitive decline.

Down-regulated transcripts include many genes involved in synaptic function (Supplemental Table 1) including synaptojanin 1 (SYNJ1) located in linkage region

21q22.2. The most significantly down-regulated transcript across all brain tissues is ATP6V1G2 (Supplemental Tables 1 and 4) located in linkage region 6p21.3. ATP6V1G2 is a membrane bound vacuolar-type ATPase that maintains the acidity of lysosomal vesicles [93]. Luminal acidification by V-ATPases is required for proper intracellular vesicle sorting and degradation of endocytosed proteins. The relationship of ATP6V1G2 to the regulation of synaptic vesicle recycling or brain aging is unknown.

In addition to appropriate retrograde transport of endosomes, synaptic plasticity is also dependent on the anterograde transport and localization of specific mRNA transcripts to the synapse. Protein synthesis occurring at the synapse is considered to be a fundamental part of healthy synaptic function. Dysregulation of microtubule subunits and molecular motors is seen in cognitive decline (Supplemental Table 1) and down-regulation of all aspects of RNA function and transport is widespread in cognitive decline (Table 4). Two transcripts related to proper mRNA localization and translation at the synapse are located in linked regions. Synaptotagmin binding, cytoplasmic RNA interacting protein (SYNCRIP, 6q14-15) is a component of mRNA granules [94] binding mRNA and ensuring proper anterograde transport [95]. SYNCRIP interacts with various isoforms of the membrane-bound synaptotagmin [96]. Molecular motor trafficking on microtubules is postulated to be blocked by protein aggregates [97]. Failure of protein aggregates to be degraded through ubiquitin-mediated proteolysis has been shown to occur in AD [98, 99] and local protein degradation through the ubiquitin-proteosome pathway has been shown to affect synaptic plasticity [100]. Many transcripts involved in this pathway are down-regulated in cognitive decline (Supplemental Table 1). A recent study suggests that cell death due to polyglutamine protein aggregates can be reduced by overexpression of RNA binding protein 3 (RBM3) [101]. RBM3 and its related gene CIRBP are down regulated in AD (Supplemental Table 1). These proteins are involved in response to stress [102]. RBM3 is located in a linkage region (Xp11.2) and has recently been shown to decrease microRNA (miRNA) levels with a parallel increase in protein synthesis [103]. MicroRNAs are small, highly conserved RNA molecules that regulate the expression of messenger RNA by binding to the 3'-untranslated regions (3'-UTR). Each miRNA is thought to regulate multiple genes and miRNA regulation is thought to influence many diverse cellular processes [104]. The contribution of miRNA regulation to cognitive decline is unknown, although miRNAs are postulated to be involved in processes related to synaptic plasticity [105].

4.2 NFT formation

NFT formation precedes cognitive decline and is correlated with severity of dementia in AD [106]. We identified a subset of genes that were differentially regulated in non-demented versus AD subjects (Hypothesis I) and subjects with low versus high tangle burden (Hypothesis II) (Supplemental Table 3). Overall, fewer transcripts were related to NFT formation and these had higher q-values than transcripts identified in the comparison of Extreme Cognitive Phenotypes (Supplemental Tables 1 and 3). This relationship is evident in other gene profiling experiments in which more transcripts were correlated with cognitive scores than NFTs [62] and more transcripts were differentially expressed in non-demented versus AD neurons than in AD non-NFT neurons versus AD NFT neurons [63]. Genes identified in this comparison may be more relevant to initial stages of brain pathology during NFT formation.

4.3 Cognitive reserve

Discovery of genes involved in individual brain capacity to tolerate, or circumvent, neuropathologic damage during aging would increase our ability to predict risk of dementia and determine risk-reducing factors. Non-demented individuals with heavy NFT burden may have more versatile neuronal processing mechanisms than individuals who develop dementia [107]. Although the limited sample size precluded statistical analyses, exploratory data analysis uncovered several genes with different patterns of expression in these subjects (Table 6).

Non-demented individuals with high Braak scores (Group 3) exhibited increased expression of a ribosomal structural gene (RPS4Y1) and the neuropeptide receptor bombesin-like receptor 3 (BRS3), compared with non-demented subjects with lower Braak scores (Groups 1 and 2) and AD subjects (Group 4). Bombesin-like neuropeptides are a family of G-protein-coupled receptors that have pleiotropic physiological effects, such as increasing hypertension and insulin secretion, stimulating gastric secretion, and modulating smooth muscle contraction [108]. Mice lacking BRS3 show mild obesity associated with hypertension, impairment of glucose tolerance and insulin resistance[109]. Dysregulated glucose metabolism has been shown to occur in AD pathology [110, 111]. Our results suggest the possibility that individual protection of brain tissue from the pathological effects of NFTs results from regulation of protein synthesis and glucose metabolism.

Of the genes that show lower expression in Group 3 subjects, one has been previously studied in AD. Glutathione S-transferase theta 1 (GSTT1) is involved in detoxification of environmental toxins, but its role in susceptibility to AD is inconclusive [112, 113]. Two genes are possibly involved in inflammatory processes. S100A8 is a subunit of Calprotectin, a calcium- and zinc-binding protein up-regulated in many inflammatory conditions [114]. Neuronal pentraxin II (NPTX2) is postulated to be involved in uptake of pro-inflammatory molecules [115]. Rat NPTX2 is regulated by synaptic activity and promotes neuronal migration [116]. Rap guanine nucleotide exchange factor 2 (RAPGEF2) is also involved in synaptic physiology through binding to a synaptic scaffold protein, and is hypothesized to link synaptic plasma membrane vesicles with RAS signal transduction [117]. These results further illustrate the central roles of anti-inflammatory processes and regulation of synaptic activity in maintaining healthy neuronal function. Additional experiments with larger sample sizes will be required to confirm the role of these genes in protection from brain tissue damage.

4.4 Genes common to the pathological process across all tissues

Determination of concordance across three transcriptomal studies allowed us to identify 174 transcripts common to cognitive decline across entorhinal cortex, hippocampus and frontal cortex. Synaptic plasticity–related genes are dysregulated in all three tissues. Likewise, down-regulation of energy pathways and ubiquitin-mediated protein degradation is widespread. Genes that function in these pathways are likely to be important in processes underlying the development of AD pathology. It is important to note that differentially expressed transcripts unique to each study may be the result of tissue specificity or non-biological differences in study design. Continued comparisons across studies and tissues will allow us to further elucidate the underlying genetic mechanisms of cognitive decline.

4.5 General considerations for transcriptomal studies

Central to the interpretation of biological significance of a particular differentially expressed transcript is the quality of the annotations obtained from publicly available databases. Often, complete annotation is not available for all of the transcripts interrogated. The annotation that does exist is dynamic and constantly updated. Finally, while it is transcripts that are interrogated on the array, it is common practice to map these transcripts to a gene index (such as Unigene ID). There can often be a loss of information in such a mapping, as it ignores differences at the transcript level. A case in point highlighted in this study is ITSN1, which is commonly found in two isoforms, a short form and a long form. Additionally, over 19 alternatively spliced forms have been identified. Affymetrix GeneChip arrays target both the short and long forms of ITSN1. In all three data sets, it was the short form only (Probe ID 35776_at) that was differentially expressed. This finding was confirmed with qRT-PCR. In the analysis of the microarray data, transcripts for the same gene are often seen as technical replicates, rather than biological variants, such that any gene with discordant ProbeSets is discarded from further analysis. This results in failure to detect unique isoforms and transcripts that may play a key role in the biological process under study.

This highlights an important aspect of the dynamic and complex nature of the annotation process that may not always be appreciated. There has been a great deal of recent debate concerning the reliability of microarray gene expression on the same samples across different platforms [118, 119]. A key point that is often missed is that in order to compare the arrays, individual transcripts are mapped to gene indices, due to the fact that different transcripts are interrogated on different platforms. There is an inherent loss of information in this mapping as alternate transcripts (each potentially with different expression patterns) are all mapped to the same gene identifier. Attempting to determine concordance based on gene annotation (such as gene symbol, name or Unigene ID) can be misleading and give overestimates of discordance, as described above.

Validation studies of microarrays using qRT-PCR also can suffer from overestimates of discordance between the arrays and the RT-PCR when primers are not designed to the same targets as the array. Strong correlations are seen between qRT-PCR and microarray results when the same transcript targets are tested [120]. This is clearly demonstrated by ITSN1 in this study, where only one transcript variant is differentially expressed, making primer design even more critical. These issues need to be considered in design of new studies and meta-analysis of existing data.

Functional genomics is often combined with whole genome association studies to improve the ability to locate susceptibility genes. We performed a pilot study using a whole genome sampling assay (WGA) to assess the characteristics of the whole genome SNP arrays and the performance of the genotyping algorithms.

Chapter II. Figures and Tables.

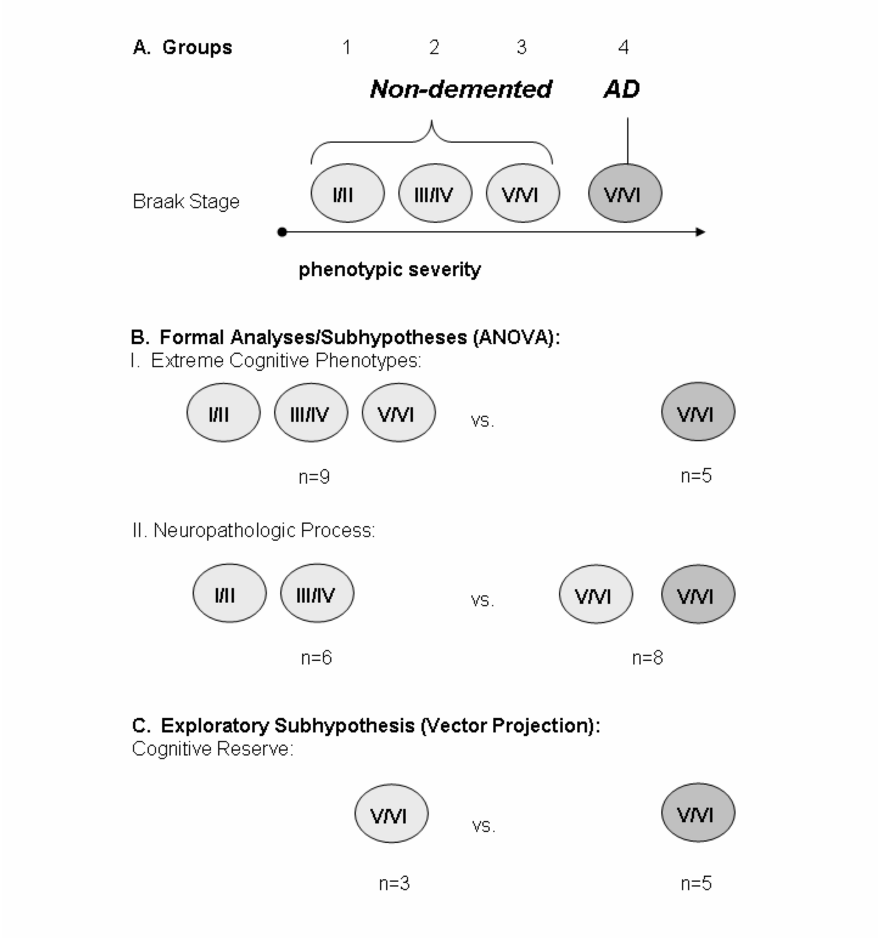


Figure 1. Subject Comparisons. A) Subjects were separated into four groups based on Braak stage and cognitive health. B) Two separate ANOVA comparisons performed. I. Extreme Cognitive Phenotypes were assessed by combining all non-demented subjects compared to AD subjects; II. Neuropathological Process was assessed by comparing low Braak stage subjects with high Braak stage subjects regardless of cognitive ability; III. Cognitive Reserve was assessed using Vector projection comparing non-demented, Braak V/VI subjects with AD subjects.

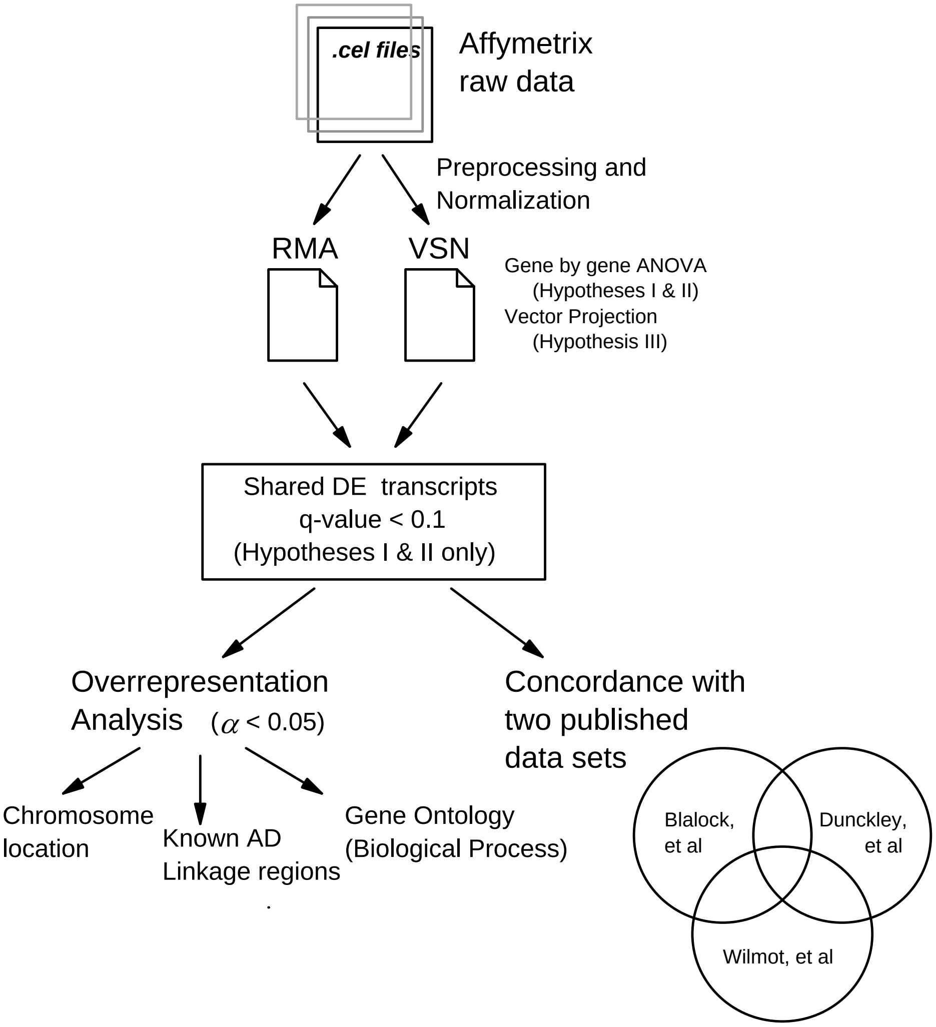
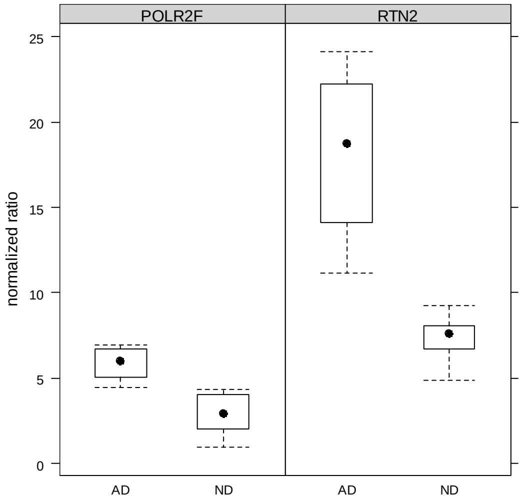

.cel files
Affymetrix raw data
Preprocessing and Normalization
RMA
VSN
Gene by gene ANOVA (Hypotheses I & II)
Vector Projection (Hypothesis III)
Shared DE transcripts q-value < 0.1 (Hypotheses I & II only)
Overrepresentation Analysis (α < 0.05)
Concordance with two published data sets
Chromosome location
Known AD Linkage regions
Gene Ontology (Biological Process)
Blalock, et al
Dunckley, et al
Wilmot, et al

Figure 2. Analytical Work Flow. Raw data files were preprocessed and normalized using two different methods. Each data set was analyzed separately by ANOVA (Hypotheses I & II) and Vector Projection (Hypothesis III). Transcripts differentially expressed (DE) in both data sets (q-value < 0.1) were combined into one data set for downstream analysis. DE transcripts were analyzed by χ^2 for overrepresentation in categories of interest (chromosome location, known AD linkage regions and Gene Ontology Biological Process categories). Concordance of DE transcripts with two previous studies was investigated.



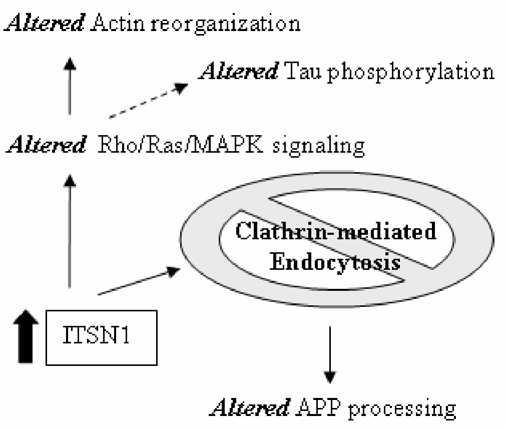

Altered Actin reorganization
Altered Tau phosphorylation
Altered Rho/Ras/MAPK signaling
Clathrin-mediated Endocytosis
ITSN1
Altered APP processing

Figure 3. Confirmation by qRT-PCR of differential expression of the short form of ITSN1 between AD and ND (all p-values < 0.05). Two reference genes, polymerase II, RNA, subunit F (POLR2F) and reticulon 2 (RTN2), were used to normalize levels of ITSN1 within AD and ND samples. Relative amounts of ITSN1 between AD and ND groups were significantly different for both reference genes, POLR2F p-value =0.009 and RTN2 p-value = 0.025.

Figure. 4. Altered metabolism due to increased expression of Intersectin1. Solid arrows are direct consequences of higher levels of ITSN1 in published reports. Dashed arrow refers to the downstream effects of the MAPK signaling cascade on the phosphorylation of Tau.

Table 1. Subject Description

	Non-demented			**AD**
Braak	**I/II**	**III/IV**	**V/VI**	**V/VI**
Age	93.03 ± 12.19	90.85 ± 0.21	93.57 ± 1.59	89.74 ± 4.33
MMSE	28.0 ± 2.0	29.0 ± 1.41	28.33 ± 1.15	14.40 ± 6.99
Clinical DX	ND	ND	ND	PRAD
	n=3	n=3	n=3	n=5

Subjects were assigned to four groups based on Braak stage scoring (see Methods). MMSE, Minimental Status Exam; Clinical DX, clinical diagnosis; ND, non-demented; Braak, Braak stage; n, number of subjects.

Table 2. Number of differentially expressed genes for each analysis[a].

	Up Regulated			**Down regulated**		
	Cognitive Differences	**NFT Formation**	**Cognitive Reserve**	**Cognitive Differences**	**NFT Formation**	**Cognitive Reserve**
Probe IDs[b]	3703	249	50	4643	279	31
Unique Genes	3664	249	48	4522	277	14
Annotated Genes	849 (23%)	41 (16%)	23 (47%)	1771 (39%)	117 (42%)	7 (50%)
GO IDs	4227	209	99	8211	520	44
Unique GO IDs	965	134	74	1544	271	41

[a] Gene Ontology (GO) Biological process category annotations for the differentially expressed genes in each analysis.

[b] Probe IDs, the number of Affymetrix Probe IDs that were differentially expressed in each analysis; Unique genes, the number of unique genes corresponding to the Probe IDs; Annotated Genes, the number of unique genes that have annotations associated with them in the GO database; GO IDs, the number of appearances of GO IDs associated with the annotated genes; Unique GO IDs, the number of unique GO IDs associated with the annotated genes. Numbers in paratheses indicate the percentage of unique genes that have associated GO annotations.

Table 3. Differentially expressed transcripts located in genomic regions linked to Alzheimer's disease[a].

	# of transcripts		
Linkage Regions	**# DE[b]**	**# on chip[c]**	**p-value[d]**
1p36	106	808	0.99
1q23-31	78	511	0.70
2p23-24	49	285	0.43
4q35	15	84	0.44
5p13-15	64	325	0.12
6p21	92	750	0.99
6q15-16	27	112	0.05
6q25-27	46	280	0.55
9p21	11	67	0.56
9q22	26	219	0.96
10q21-22	55	319	0.42
10q25	22	120	0.37
12p11-12	30	219	0.86
19q13	144	1256	1.00
21q21-22	66	499	0.96
Xp11-21	42	321	0.94

[a] Linkage regions are reproduced from Bertram and Tanzi {Bertram, 2004 #72}. Transcripts differentially expressed (DE) between non-demented and AD subjects (hypothesis I) were compared for overrepresentation in linkage regions.

[b] DE, number of transcripts differentially expressed at $q < 0.1$ by ANOVA that are located in the linkage region

[c] Number of transcripts on the chip that are located in the linkage region

[d] p-values are from Fisher's Exact test comparing transcripts DE at $q < 0.1$ to all transcripts on the Affymetrix HGU133Plus2 GeneChip in each linkage region.

Table 4. Biological process categories significantly overrepresented in Cognitive Differences Hypothesis (non-demented vs. AD)

A. up regulated in AD

Category	# genes DE[a]	# genes on chip[b]	FDR[c]
regulation of cellular physiological process	208	2252	3.21E-05
regulation of biological process	231	2578	3.21E-05
regulation of cellular process	215	2365	3.21E-05
regulation of physiological process	214	2376	6.34E-05
regulation of transcription, DNA dependent	149	1533	6.35E-05
regulation of transcription	154	1598	6.35E-05
regulation of nucleic acid metabolism	155	1620	8.39E-05
regulation of cellular metabolism	157	1650	8.94E-05
transcription	159	1684	0.000115
transcription, DNA-dependent	151	1587	0.000137
regulation of metabolism	159	1733	0.000752
negative regulation of cellular physiological process	48	384	0.0011
actin filament-based process	17	88	0.00146
negative regulation of physiological process	49	412	0.00451
negative regulation of cellular process	49	416	0.0059
chromatin modification	16	87	0.00717
actin cytoskeleton organization and biogenesis	16	77	0.0114
negative regulation of biological process	51	456	0.0189
phosphate transport	14	78	0.0363
nucleic acid metabolism	205	2502	0.0418

B. Down regulated in AD

Category	# genes	# genes on chip	FDR
coenzyme metabolism	65	124	1.69E-32
cofactor metabolism	70	145	1.42E-30
oxidative phosphorylation	41	65	2.47E-27
coenzyme biosynthesis	41	81	1.25E-18
cofactor biosynthesis	45	96	6.15E-18
biosynthesis	217	919	3.27E-17
ribonucleotide biosynthesis	34	69	1.76E-14
nucleoside phosphate metabolism	25	42	1.76E-14
ATP biosynthesis	25	42	1.76E-14
ATP coupled proton transport	23	38	1.30E-13
energy coupled proton transport, down electrochemical gradient	23	38	1.30E-13
ribonucleotide metabolism	34	72	1.58E-13
ATP metabolism	25	44	1.76E-13
group transfer coenzyme metabolism	28	54	4.57E-13
ribonucleotide triphosphate biosynthesis	27	51	4.57E-13
purine ribonucleotide triphosphate biosynthesis	27	51	4.57E-13
purine nucleoside triphosphate biosynthesis	27	51	4.57E-13
nucleoside triphosphate metabolism	27	52	1.27E-12
generation of precursor metabolites and energy	128	504	1.33E-12

intracellular transport	114	433	1.63E-12
cellular biosynthesis	183	807	2.29E-12
ribonucleoside triphosphate metabolism	27	53	2.55E-12
purine ribonucleoside triphosphate metabolism	27	53	2.55E-12
purine nucleoside triphosphate metabolism	27	53	2.55E-12
purine ribonucleotide biosynthesis	30	64	7.52E-12
purine nucleotide biosynthesis	31	68	1.18E-11
nucleoside triphosphate metabolism	27	55	1.59E-11
establishment of protein localization	108	417	2.77E-11
purine ribonucleotide metabolism	30	66	3.22E-11
purine nucleotide metabolism	31	70	4.56E-11
purine nucleotide metabolism	107	415	4.80E-11
protein transport	42	113	1.10E-10
nucleotide biosynthesis	108	425	1.23E-10
hydrogen transport	29	65	1.86E-10
proton transport	28	64	1.19E-09
nucleotide metabolism	51	159	2.90E-09
intracellular protein transport	74	271	7.39E-09
aerobic respiration	18	27	5.76E-08
cellular respiration	18	29	3.12E-07
ATP synthesis coupled electron transport	16	25	1.28E-06
metabolism	933	6083	1.75E-06
translation	42	138	1.75E-06
RNA metabolism	82	340	1.75E-06
ATP synthesis coupled electron transport	15	23	2.41E-06
acety-CoA metabolism	15	24	5.36E-06
main pathways of carbohydrate metabolism	29	84	5.82E-06
energy derivation by oxidation of organic compounds	38	127	1.61E-05
tricarboxylic acid cycle	13	20	2.10E-05
acety-CoA catabolism	13	20	2.10E-05
coenzyme catabolism	13	20	2.10E-05
cofactor catabolism	14	23	2.23E-05
secretory pathway	35	116	3.75E-05
cellular metabolism	869	5712	6.16E-05
macromolecule metabolism	462	2830	9.62E-05
RNA processing	63	266	0.000178
secretion	40	146	0.000178
protein folding	49	192	0.000189
mitochondrial electron transport, NADH to ubiquinone	12	20	0.000191
mRNA metabolism	44	167	0.000211
cellular macromolecule metabolism	436	2674	0.000251
cellular physiological process	1164	7955	0.000273
protein biosynthesis	95	455	0.000434
RNA splicing, via transesterif	26	83	0.000452
nuclear mRNA splicing, via spliceosome	26	83	0.000452
RNA splicing, via transesterif	26	83	0.000452
electron transport	70	312	0.000476
establishment of localization	305	1803	0.000597

mRNA processing	39	147	0.000626
transport	304	1801	0.000739
localization	305	1810	0.000823
RNA splicing	31	111	0.00151
protein targeting	31	111	0.00151
biopolymer metabolism	235	1361	0.00199
macromolecule biosynthesis	102	513	0.00205
Golgi vesicle transport	16	44	0.00229
protein-mitochondrial targeting	10	18	0.00292
sterol biosynthesis	12	25	0.00316
primary metabolism	815	5466	0.00606
protein metabolism	400	2514	0.00833
mitochondrial organization and biogenesis	8	14	0.0141
translational initiation	16	49	0.0175
cellular protein metabolism	394	2496	0.0193
inner mitochondrial membrane organization and biogenesis	5	6	0.0206
mitochondrial inner membrane protein import	5	6	0.0206
regulated secretory pathway	8	15	0.0248
quinone cofactor metabolism	4	4	0.0275
quinone cofactor biosynthesis	4	4	0.0275
ubiquinone biosynthesis	4	4	0.0275
ubiquinone metabolism	4	4	0.0275
lipid biosynthesis	39	170	0.036

[a] the number of differentially expressed genes (q< 0.1 by ANOVA) that are members of the category

[b] the number of genes on the Affymetrix GeneChip that are members of the category

[c] the FDR values are from χ^2 analysis corrected for multiple testing (see methods)

Table 5. Canonical Pathways involved in healthy aging

Pathway	**# genes**[a]	**# genes in pathway**[b]	**p-value**[c]	**% A**[d]
Oxidative phosphorylation	66	130	0.000	50.769
Proteasome	20	31	0.000	64.516
ATP synthesis	22	40	0.000	55.000
Infection	18	41	0.009	43.902
Citrate cycle (TCA cycle)	12	25	0.014	48.000
Synthesis and degradation of ketone bodies	5	7	0.015	71.429
RNA polymerase	11	23	0.019	47.826
Carbon fixation	10	22	0.037	45.455
Phenylalanine	6	11	0.041	54.545
Butanoate metabolism	17	44	0.044	38.636
Amyotrophic lateral sclerosis (ALS)	8	17	0.049	47.059
Reductive carboxylate cycle (CO2 fixation)	5	9	0.056	55.556
Valine	16	43	0.068	37.209
Aminoacyl-tRNA biosynthesis	10	25	0.088	40.000

[a] the number of differentially expressed genes (q < 0.1 by ANOVA) that are members of the pathway

[b] the total number of genes in the pathway

[c] p-values are from a one-tailed Fisher's Exact Test (see methods)

[d] % abundance of differentially expressed genes in that pathway

Table 6. Transcripts with maximum Differences between Group 3 (non-demented with high Braak score) and other Groups.

A. Transcripts with an increased transcript in non-demented, high Braak stage subjects.

Probe ID	Gene	Symbol	chromosome location
207369_at	bombesin-like receptor 3	BRS3	Xq26-q28
226558_at	hypothetical gene supported by AK096952; AK126241; BC068588	LOC441057	4p16.3
238774_at	Hypothetical protein LOC284058	LOC284058	17q21.31
201909_at	ribosomal protein S4, Y-linked 1	RPS4Y1	Yp11.3
B. Transcripts with a decreased transcript in non-demented, high Braak stage subjects.			
203815_at	glutathione S-transferase theta 1	GSTT1	22q11.23
224588_at			
227671_at			
203096_s_at	Rap guanine nucleotide exchange factor (GEF) 2	RAPGEF2	4q32.1
221728_x_at			
202917_s_at	S100 calcium binding protein A8 (calgranulin A)	S100A8	1q21
213479_at	neuronal pentraxin II	NPTX2	7q21.3-q22.1

Table 7. Concordance Rates per data set for Affymetrix HG_U133A GeneChip array.[a]

A. Intersection using Probe ID

	10% FDR		5% FDR	
	UP	**DN**	**UP**	**DN**
Blalock[b]	104 (8.4%)	493 (13.7%)	55 (9.4%)	392 (16.7%)
Dunckley	252 (20.4%)	733 (20.4%)	117 (19.9%)	488 (20.8%)
Blalock/Dunckley	451 (10.1%)	239 (8.4%)	254 (7.1%)	141 (10.0%)
all 3 data sets	33 (2.6%)	141 (3.9%)	6 (1.0%)	54 (2.3%)

B. Intersection using Gene Symbol

	10% FDR	
	UP	**DN**
Blalock[b]	258 (35.5%)	610 (27.5%)
Dunckley	195 (26.8%)	812 (36.6%)
Blalock/Dunckley	762 (21.7%)	554 (14.0%)
all 3 data sets	59 (8.1%)	274 (12.3%)

[a] Concordance was determined for each pair of data sets by measuring the intersection of transcripts significantly differentially regulated at 10% and 5% FDR. The number and percentage of transcripts concordant in each comparison is given.

[b] The concordance of each data set with our results, between Blalock, et al and Dunckley, et al, and the concordance among all three data sets is presented.

Table 8. Transcripts differentially expressed in non-demented versus demented that are common to all data sets.[a]

Upregulated in AD

Probe ID	Symbol	chromosome location[b]	p-value Blalock et al	p-value Dunckley et al
35776_at	ITSN1	**21q22.1-q22.2**	0.05	0.00
201502_s_at	NFKBIA	14q13	0.01	0.00
202273_at	PDGFRB	5q31-q32	0.07	0.00
201125_s_at	ITGB5	3q21.2	0.00	0.02
202861_at	PER1	17p13.1-17p12	0.04	0.00
210473_s_at	GPR125	4p15.31	0.07	0.01
203685_at	BCL2	18q21.33, 18q21.3	0.00	0.07
210069_at	CHKB /// CPT1B	22q13.33	0.02	0.06
206766_at	ITGA10	1q21	0.08	0.00
221527_s_at	PARD3	10p11.22-p11.21	0.02	0.04
212346_s_at	MXD4	4p16.3	0.08	0.02
213044_at	ROCK1	18q11.1	0.10	0.03
203505_at	ABCA1	9q31.1	0.01	0.01
36829_at	PER1	17p13.1-17p12	0.03	0.00
217937_s_at	HDAC7A	12q13.1	0.08	0.00
214594_x_at	ATP8B1	18q21-q22, 18q21.31	0.06	0.00
209703_x_at	DKFZP586A0522	12q13.12	0.04	0.00
205168_at	DDR2	**1q12-q23**	0.00	0.04
221191_at	DKFZP434A0131	7q11.23-q21.1	0.02	0.02
203080_s_at	BAZ2B	2q23-q24	0.07	0.04
204060_s_at	PRKX /// PRKY	Xp22.3 , Yp11.2	0.00	0.01
212122_at	RHOQ /// LOC284988	2p21 , 2q21.1	0.00	0.03
209370_s_at	SH3BP2	4p16.3	0.02	0.06
202724_s_at	FOXO1A	13q14.1	0.00	0.01
205111_s_at	PLCE1	10q23	0.03	0.00
205288_at	CDC14A	1p21	0.03	0.00
204061_at	PRKX	Xp22.3	0.02	0.00
202933_s_at	YES1	18p11.31-p11.21	0.02	0.03
209108_at	TM4SF6	Xq22	0.01	0.00

Downregulated in AD

Probe ID	Symbol	chromosome location	p-value Blalock et al	p-value Dunckley et al
214762_at	ATP6V1G2	**6p21.3**	0.01	0.02
221020_s_at	MFTC	8q22.3	0.06	0.00
210976_s_at	PFKM	12q13.3	0.02	0.04
219443_at	C20orf13	20p12.1	0.07	0.01
203889_at	SGNE1	15q13-q14	0.05	0.09
202325_s_at	ATP5J	**21q21.1**	0.09	0.03
201304_at	NDUFA5	7q32	0.01	0.09
204675_at	SRD5A1	5p15	0.08	0.00

222005_s_at	GNG3	11p11	0.03	0.00
200720_s_at	ACTR1A	10q24.32	0.06	0.00
208934_s_at	LGALS8	1q42-q43	0.02	0.01
218291_at	MAPBPIP	1q22	0.02	0.06
206290_s_at	RGS7	1q43	0.01	0.00
206489_s_at	DLGAP1	18p11.3	0.03	0.01
218488_at	EIF2B3	1p34.1	0.03	0.02
213849_s_at	PPP2R2B	5q31-5q32	0.01	0.02
215161_at	CAMK1G	1q32-q41	0.00	0.04
204471_at	GAP43	3q13.1-q13.2	0.03	0.00
200039_s_at	PSMB2	1p34.2	0.03	0.01
213011_s_at	TPI1	12p13	0.02	0.02
206055_s_at	SNRPA1	15q26.3	0.07	0.09
209583_s_at	CD200	3q12-q13	0.03	0.08
203218_at	MAPK9	5q35	0.09	0.09
211023_at	PDHB	3p21.1-p14.2	0.00	0.02
210027_s_at	APEX1	14q11.2-q12	0.02	0.03
221471_at	TDE1	20q13.1-13.3	0.02	0.01
218332_at	BEX1	Xq21-q23	0.06	0.00
213666_at	6-Sep	Xq24	0.00	0.03
210014_x_at	IDH3B	20p13	0.04	0.03
201569_s_at	CGI-51	22q13.31	0.07	0.00
211276_at	TCEAL2	Xq22.1-q22.3	0.02	0.00
202634_at	POLR2K	8q22.2	0.02	0.01
207142_at	KCNJ3	2q24.1	0.10	0.08
221482_s_at	ARPP-19	15q21.2	0.05	0.06
206342_x_at	IDS	Xq28	0.10	0.00
200822_x_at	TPI1	12p13	0.04	0.03
212990_at	SYNJ1	**21q22.2**	0.06	0.02
208870_x_at	ATP5C1	10p15.1	0.03	0.08
200613_at	AP2M1	3q28	0.03	0.00
218193_s_at	GOLT1B	**12p12.1**	0.08	0.00
217948_at	DKFZP564B147	Xq26.3	0.02	0.02
202961_s_at	ATP5J2	7q22.1	0.10	0.02
202279_at	C14orf2	14q32.33	0.01	0.00
218404_at	SNX10	7p15.2	0.02	0.02
204744_s_at	IARS	9q21	0.05	0.04
202596_at	ENSA	1q21.2	0.00	0.09
209075_s_at	NIFUN	12q24.1	0.03	0.08
205549_at	PCP4	**21q22.2**	0.01	0.01
218813_s_at	SH3GLB2	9q34	0.07	0.01
208308_s_at	GPI	**19q13.1**	0.09	0.05
208745_at	ATP5L	11q23.3	0.04	0.04
200001_at	CAPNS1	**19q13.12**	0.10	0.07
208906_at	BSCL2	11q12-q13.5	0.08	0.08
206089_at	NELL1	11p15.2-p15.1	0.00	0.08
205711_x_at	ATP5C1	10p15.1	0.01	0.02
219196_at	SCG3	15q21	0.04	0.05
209025_s_at	SYNCRIP	**6q14-q15**	0.00	0.00
212826_s_at	SLC25A6	Xp22.32 and Yp	0.01	0.01

209482_at	POP7	7q22	0.04	0.08
211698_at	CRI1	15q21.1-q21.2	0.01	0.10
201849_at	BNIP3	10q26.3	0.03	0.00
201797_s_at	VARS2	**6p21.3**	0.04	0.07
205899_at	CCNA1	13q12.3-q13	0.08	0.07
201662_s_at	ACSL3	2q34-q35	0.04	0.02
209056_s_at	CDC5L	**6p21**	0.01	0.02
201524_x_at	UBE2N	12q22	0.01	0.04
219619_at	DIRAS2	**9q22.2**	0.04	0.06
206233_at	B4GALT6	18q11	0.05	0.00
213333_at	MDH2	7p12.3-q11.2	0.02	0.09
203079_s_at	CUL2	10p11.21	0.02	0.03
213902_at	ASAH1	8p22-p21.3	0.04	0.01
201400_at	PSMB3	17q12	0.02	0.01
215691_x_at	C1orf41	1p32.1-p33	0.01	0.00
217773_s_at	NDUFA4	7p21.3	0.01	0.05
201568_at	QP-C	5q31.1	0.06	0.02
206857_s_at	FKBP1B	**2p23.3**	0.08	0.00
214436_at	FBXL2	3p23	0.03	0.01
208977_x_at	TUBB2	6p25	0.03	0.06
200650_s_at	LDHA	11p15.4	0.01	0.00
216120_s_at	ATP2B2	3p25.3	0.01	0.00
212296_at	PSMD14	2q24.2	0.07	0.00
221437_s_at	MRPS15	1p35-p34.1	0.09	0.01
201381_x_at	CACYBP	**1q24-q25**	0.07	0.00
206381_at	SCN2A2	2q23-q24	0.09	0.04
209849_s_at	RAD51C	17q22-q23	0.00	0.03
206949_s_at	RUSC1	1q21-q22	0.03	0.00
208975_s_at	KPNB1	17q21.32	0.04	0.08
210406_s_at	RAB6A /// RAB6C	11q13.3 , 2q31	0.01	0.01
200027_at	NARS	18q21.2-q21.3	0.03	0.00
209914_s_at	NRXN1	2p16.3	0.08	0.02
214005_at	GGCX	2p12	0.06	0.00
201597_at	COX7A2	6q12	0.05	0.00
213366_x_at	ATP5C1	10p15.1	0.01	0.02
219073_s_at	OSBPL10	3p22.3	0.00	0.00
208905_at	CYCS	7p15.3	0.03	0.00
217801_at	ATP5E	20q13.32	0.03	0.05
202309_at	MTHFD1	14q24	0.00	0.00
203894_at	TUBG2	17q21	0.00	0.02
209877_at	SNCG	10q23.2-q23.3	0.02	0.09
216903_s_at	CBARA1	**10q22.1**	0.09	0.01
202260_s_at	STXBP1	9q34.1	0.08	0.00
201837_s_at	STAF65(gamma)	2pter-p25.1	0.00	0.00
218226_s_at	NDUFB4	3q13.33	0.06	0.00
207081_s_at	PIK4CA	22q11.21	0.02	0.07
209142_s_at	UBE2G1	1q42, 17p13.2	0.07	0.00
220045_at	NEUROD6	7p14.3	0.01	0.02
202090_s_at	UQCR	19p13.3	0.03	0.00
200734_s_at	ARF3	12q13	0.01	0.00

213726_x_at	TUBB2	6p25	0.02	0.04
201047_x_at	RAB6A	11q13.3	0.06	0.07
204141_at	TUBB2	6p25	0.02	0.00
210016_at	MYT1L	2p25.3	0.01	0.03
208936_x_at	LGALS8	1q42-q43	0.01	0.00
205691_at	SYNGR3	16p13	0.01	0.00
203001_s_at	STMN2	8q21.13	0.09	0.00
218732_at	Bit1	17q23.2	0.10	0.05
205113_at	NEF3	8p21	0.01	0.05
218106_s_at	MRPS10	**6p21.1-p12.1**	0.06	0.05
203846_at	TRIM32	9q33.1	0.01	0.00
209001_s_at	ANAPC13	3q22.1	0.02	0.00
203797_at	VSNL1	**2p24.3**	0.01	0.07
203303_at	TCTE1L	**Xp21**	0.05	0.00
211071_s_at	AF1Q	1q21	0.09	0.00
204247_s_at	CDK5	7q36	0.01	0.04
221288_at	GPR22	7q22-q31.1	0.01	0.07
201434_at	TTC1	5q32-q33.2	0.03	0.00
212976_at	TA-LRRP	1p22.2	0.03	0.07
203667_at	TBCA	5q14.1	0.01	0.05
200625_s_at	CAP1	1p34.2	0.02	0.00
218467_at	TNFSF5IP1	18p11.21	0.02	0.01
204465_s_at	INA	10q24.33	0.01	0.00
202754_at	R3HDM	2q21.3	0.02	0.01
215518_at	STXBP5L	3q13.33	0.05	0.00
222125_s_at	PH-4	3p21.31	0.02	0.02
206051_at	ELAVL4	1p34	0.09	0.03
202336_s_at	PAM	5q14-q21	0.10	0.00
202022_at	ALDOC	17cen-q12	0.02	0.08
201980_s_at	RSU1	10p13	0.01	0.01
211069_s_at	SUMO1	2q33	0.05	0.08
201527_at	ATP6V1F	7q32	0.06	0.05

[a] Differentially expressed transcripts are $q < 0.1$ from ANOVA

[b] Values in bold are chromosomal regions linked to AD

Supplemental Table 1. Transcripts differentially expressed between non-demented and AD subjects (Hypothesis 1).

Up regulated in AD

Probe ID	symbol	chromosome location	rma qvalues	vsn qvalues	linked
35776_at	ITSN1	21q22.1-q22.2	0.006357	0.002705	Y
1555842_at	LOC284356	19q13.33	0.006357	0.004299	Y
232002_at	GPI	19q13.1	0.006554	0.004299	Y
239449_at	ANKH	5p15.1	0.006357	0.004331	Y
235705_at	TRIO	5p15.1-p14	0.006357	0.00438	Y
212177_at	C6orf111	6q16.3	0.006357	0.004873	Y
233816_at	SLC8A1	2p23-p22	0.00778	0.005188	Y
244530_at	FLJ25422	5p13.2	0.006554	0.005268	Y
228028_at	LOC150946	2p23.3	0.007969	0.005364	Y
232621_at	USP48	1p36.12	0.006357	0.005449	Y
242022_at	PBX1	1q23	0.011337	0.006493	Y
243826_at	NCOA1	2p23	0.007114	0.006511	Y
244015_at	EIF4G3	1p36.12	0.011625	0.006528	Y
240008_at	ARID1B	6q25.1	0.008594	0.007535	Y
233041_x_at	BTBD9	6p21	0.008819	0.007667	Y
223915_at	BCOR	Xp11.4	0.010728	0.007673	Y
225507_at	C6orf111	6q16.3	0.009534	0.008666	Y
233025_at	PDZK3	5p13.3	0.010849	0.008751	Y
231403_at	TRIO	5p15.1-p14	0.009366	0.00898	Y
232592_at	RPL10A	6p21.3-p21.2	0.015604	0.009295	Y
227901_at	FLJ90723	5p13.1	0.017569	0.00957	Y
240205_x_at	KIAA0528	12p12.1	0.010782	0.009812	Y
244753_at	ACTN4	19q13	0.012393	0.009812	Y
242578_x_at	SLC22A3	6q26-q27	0.009218	0.009812	Y
232601_at		10q22.3	0.013874	0.010076	Y
241837_at	ARID5B	10q21.2	0.013874	0.010231	Y
209297_at	ITSN1	21q22.1-q22.2	0.009065	0.010843	Y
229145_at	C10orf104	10q22.1	0.012807	0.0111	Y
238672_at	PEX6	6p21.1	0.017146	0.01172	Y
212179_at	C6orf111	6q16.3	0.009903	0.011901	Y
229957_at	BCKDHA	19q13.1-q13.2	0.015789	0.012059	Y
213600_at	SIPA1L3	19q13.13	0.014215	0.012189	Y
211841_s_at	TNFRSF25	1p36.2	0.01602	0.012383	Y
236354_at	ZMYND17	10q22.2	0.012511	0.012425	Y
224667_x_at	C10orf104	10q22.1	0.018266	0.012656	Y
1554595_at	SYMPK	19q13.3	0.025039	0.012715	Y
230572_at	FLJ30277	4q35.1	0.008594	0.012749	Y
213647_at	DNA2L	10q21.3-q22.1	0.012399	0.013196	Y
234041_at	FLJ20054	1q31.3	0.017089	0.013238	Y
215109_at	KIAA0492	1q25.1	0.014215	0.013266	Y
238009_at	SOX5	12p12.1	0.013658	0.013537	Y

1569519_at	LOC439928	1p36.13	0.011156	0.013557	Y
241509_at	PLEKHA5	12p12	0.015277	0.013564	Y
215907_at	BACH2	6q15	0.017685	0.013653	Y
243766_s_at	TEAD2	19q13.3	0.02424	0.013703	Y
230368_at	ERF	19q13	0.018107	0.013879	Y
214989_x_at	PLEKHA5	12p12	0.008861	0.013891	Y
233291_at	ODZ3	4q35.1	0.016176	0.014175	Y
210847_x_at	TNFRSF25	1p36.2	0.021676	0.014837	Y
211920_at	BF	6p21.3	0.020535	0.015265	Y
222310_at	SFRS15	21q22.1	0.014903	0.016067	Y

Downregulated in AD

214762_at	ATP6V1G2	6p21.3	0.006357	0.002268	Y
212453_at	KIAA1279	10q22.1	0.006554	0.00259	Y
225936_at	CRI2	19q13.2	0.006357	0.002961	Y
229233_at	NRG3	10q22-q23	0.006696	0.002976	Y
203798_s_at	VSNL1	2p24.3	0.006554	0.002976	Y
218573_at	MAGEH1	Xp11.22	0.008095	0.003081	Y
208675_s_at	DDOST	1p36.1	0.006554	0.003759	Y
1557915_s_at	GSTO1	10q25.1	0.006765	0.003823	Y
214512_s_at	PC4	5p13.3	0.006554	0.003823	Y
208682_s_at	MAGED2	Xp11.2	0.009387	0.004047	Y
227669_at	DKFZP564B167	1q24	0.006554	0.004299	Y
1554351_a_at	MGC3794	1q23.2	0.006554	0.004299	Y
209326_at	SLC35A2	Xp11.23-p11.22	0.007351	0.004299	Y
205052_at	AUH	9q22.31	0.009525	0.004299	Y
218111_s_at	CMAS	12p12.1	0.006554	0.004299	Y
202675_at	SDHB	1p36.1-p35	0.011337	0.004873	Y
203405_at	DSCR2	21q22.3	0.008133	0.004991	Y
203613_s_at	NDUFB6	9p21.1	0.010071	0.005188	Y
209755_at	NMNAT2	1q25	0.007565	0.005188	Y
223123_s_at	HT014	1p36.11	0.006357	0.005188	Y
202325_s_at	ATP5J	21q21.1	0.013137	0.005192	Y
211662_s_at	VDAC2	10q22	0.016572	0.005207	Y
222991_s_at	UBQLN1	9q22, 9q21.2-q21.3	0.009411	0.005218	Y
200886_s_at	PGAM1	10q25.3	0.013279	0.00534	Y
204675_at	SRD5A1	5p15	0.006554	0.005355	Y
218824_at	FLJ10781	19q13.32	0.010396	0.005449	Y
218597_s_at	C10orf70	10q21.1	0.010729	0.005449	Y
228009_x_at	ZNRD1	6p21.3	0.006554	0.005449	Y
210232_at	CDC42	1p36.1	0.006554	0.005566	Y
225446_at	WDR9	21q22.2	0.006554	0.005575	Y
218682_s_at	SLC4A1AP	2p23.3-p23.2	0.010767	0.005724	Y
204230_s_at	SLC17A7	19q13	0.014215	0.005724	Y
212857_x_at	PC4	5p13.3	0.02757	0.005724	Y
215167_at	CRSP2	Xp11.4-p11.2	0.009541	0.005724	Y
224587_at	PC4	5p13.3	0.01091	0.005724	Y
222436_s_at	VPS24	2p24.3-p24.1	0.014671	0.005724	Y
202427_s_at	DKFZP564B167	1q24	0.013215	0.006048	Y
200820_at	PSMD8	19q13.2	0.011601	0.006055	Y

200620_at	C1orf8	1p36-p31	0.0298	0.006086	Y
223247_at	MGC5309	5p15.31	0.014891	0.006216	Y
224586_x_at	PC4	5p13.3	0.018107	0.006233	Y
209158_s_at	PSCD2	19q13.3	0.010921	0.006233	Y
1558700_s_at	LOC339324	19q13.12	0.008323	0.006286	Y
1558201_s_at	SLC4A1AP	2p23.3-p23.2	0.009501	0.006405	Y
205355_at	ACADSB	10q25-q26	0.008133	0.006493	Y
205540_s_at	RRAGB	Xp11.22	0.006357	0.006493	Y
218312_s_at	FLJ12895	19q13.43	0.023917	0.006499	Y
200622_x_at	CALM3	19q13.2-q13.3	0.017856	0.006511	Y
200623_s_at	CALM3	19q13.2-q13.3	0.013941	0.006663	Y
218545_at	FLJ11088	12p11.22	0.010729	0.006678	Y
238487_at	LOC285831	6p21.32	0.009349	0.006904	Y

The top 50 upregulated and top 50 downregulated transcripts are shown. For the complete list of transcripts, see [ ].

Supplemental Table 2. Cytogenetic bands containing an overrepresentation of differentially expressed transcripts.[a]

	# of transcripts		
cytogenetic bands	# DE[b]	# on chip[c]	p-value[d]
2p15	14	44	0.023
3p24	30	102	0.007
3p25	48	191	0.013
3q25	37	143	0.02
5q12	32	111	0.006
6p25	29	116	0.043
8q22	45	161	0.003
10p13	26	80	0.004
18p11	46	191	0.025
20p12	32	94	0

[a] Transcripts were tested for the number of differentially expressed transcripts that were overrepresented in a particular cytogenetic band by χ^2 analysis.

[b] DE, number of differentially expressed transcripts at $q < 0.1$ by ANOVA located in the cytogenetic band

[c] Number of transcripts on the Affymetrix HGU133Plus2 GeneChip that are located in the cytogenetic band

[d] p-values are from χ^2 analysis and adjusted by simulation

Supplemental Table 3. Differentially expressed transcripts between low and high NFT pathology (Hypothesis 2).

Upregulated in high Braak					
Probe ID	**Symbol**	**chromosome location**	**rma qvalues**	**vsn qvalues**	**linked**
239449_at	ANKH	5p15.1	0.08326	0.0524	Y
232621_at	USP48	1p36.12	0.08326	0.0524	Y
235705_at	TRIO	5p15.1-p14	0.08326	0.0524	Y
215109_at	KIAA0492	1q25.1	0.08326	0.0524	Y
242188_at	RPL10A	6p21.3-p21.2	0.08326	0.0524	Y
240008_at	ARID1B	6q25.1	0.08326	0.0533	Y
232002_at	GPI	19q13.1	0.08976	0.05474	Y
233816_at	SLC8A1	2p23-p22	0.09738	0.05534	Y
244015_at	EIF4G3	1p36.12	0.09738	0.06065	Y
244530_at	FLJ25422	5p13.2	0.09621	0.0654	Y
231403_at	TRIO	5p15.1-p14	0.09947	0.07305	Y
243826_at	NCOA1	2p23	0.09921	0.07356	Y
212179_at	C6orf111	6q16.3	0.08788	0.07379	Y
235788_at	TRIM26	6p21.3	0.08788	0.07527	Y
234041_at	FLJ20054	1q31.3	0.09552	0.07787	Y
241041_at	SLC8A1	2p23-p22	0.09958	0.07787	Y
212177_at	C6orf111	6q16.3	0.08788	0.07787	Y
220348_at	KBTBD9	2p24.1	0.09621	0.08295	Y
216628_at	ITSN1	21q22.1-q22.2	0.09947	0.08804	Y
243715_at	RNF138P1	5q11.2	0.08326	0.0479	N
230663_at	FMNL2	2q23.3	0.08326	0.0479	N
232286_at	LAF4	2q11.2-q12	0.08326	0.0479	N
237881_at	IGF1R	15q26.3	0.08326	0.0479	N
1566772_at	SCHIP1	3q25.32-q25.33	0.08326	0.0479	N
234628_at	RAB28	4p15.33	0.08326	0.0479	N
233648_at	SIK2	11q23.1	0.08326	0.0479	N
232606_at	ANK2	4q25-q27	0.08326	0.0479	N
1560924_at			0.08326	0.0479	N
1560082_at			0.08326	0.0479	N
1557745_at			0.08326	0.0479	N
224098_at			0.08326	0.0479	N
214656_x_at	MYO1C	17p13	0.08326	0.0479	N
242865_at	SDFR1	15q22	0.08326	0.0479	N
233056_x_at	DLGAP4	20q11.23	0.0921	0.05053	N
230791_at	NFIB	9p24.1	0.08326	0.05114	N
215591_at	SATB2	2q33	0.08326	0.0524	N
232599_at	SEC15L1	10q23.33	0.08326	0.0524	N
236610_at	PDE4D	5q12	0.08326	0.0524	N
244605_at			0.08326	0.0524	N
225342_at	AK3	9pter-p13	0.08326	0.0524	N
224549_x_at			0.08326	0.0524	N
244697_at	ZBTB16	11q23.1	0.08788	0.0524	N
241786_at	PPP3R1	2p15	0.08326	0.0524	N
1560865_a_at			0.08326	0.0524	N
244055_at			0.08326	0.0524	N
233219_at	MKLN1	7q32	0.08326	0.0524	N
242235_x_at	NRD1	1p32.2-p32.1	0.08326	0.0524	N
1559820_at	APG10L	5q14.1-q14.2	0.08326	0.0524	N
237174_at			0.09059	0.0524	N
232935_at	LHFP	13q12	0.08326	0.0524	N

238883_at	THRAP2	12q24.21	0.08326	0.0524	N
Downregulated in high Braak					
223247_at	MGC5309	5p15.31	0.08326	0.0479	Y
1557915_s_at	GSTO1	10q25.1	0.08326	0.0479	Y
214762_at	ATP6V1G2	6p21.3	0.08326	0.0479	Y
225936_at	CRI2	19q13.2	0.08326	0.04874	Y
1555935_s_at	HUNK	21q22.1	0.08326	0.0524	Y
218597_s_at	C10orf70	10q21.1	0.09621	0.0524	Y
218111_s_at	CMAS	12p12.1	0.08788	0.0524	Y
228009_x_at	ZNRD1	6p21.3	0.08326	0.0524	Y
219219_at	FLJ20512	19q13.32	0.08788	0.0524	Y
214512_s_at	PC4	5p13.3	0.09621	0.0524	Y
204675_at	SRD5A1	5p15	0.08326	0.0524	Y
227669_at	DKFZP564B167	1q24	0.09621	0.0524	Y
202675_at	SDHB	1p36.1-p35	0.09621	0.0524	Y
226896_at	CHCHD1	10q22.2	0.08326	0.0524	Y
222991_s_at	UBQLN1	9q22, 9q21.2-q21.3	0.09738	0.05272	Y
205549_at	PCP4	21q22.2	0.08326	0.05465	Y
1554351_a_at	MGC3794	1q23.2	0.09327	0.05534	Y
219693_at	AGPAT4	6q26	0.09621	0.05545	Y
219531_at	Cep72	5p15.33	0.08326	0.05618	Y
210959_s_at	SRD5A1	5p15	0.08326	0.05753	Y
218545_at	FLJ11088	12p11.22	0.08915	0.05834	Y
204331_s_at	MRPS12	19q13.1-q13.2	0.08915	0.06761	Y
1554133_at	RUFY2	10q21.3	0.08326	0.06836	Y
211704_s_at	SPIN2	Xp11.1	0.08906	0.06926	Y
229289_at	LOC284361	19q13.33	0.08788	0.07158	Y
201490_s_at	PPIF	10q22-q23	0.09621	0.07163	Y
218453_s_at	C6orf35	6q25.3	0.09657	0.07329	Y
231739_at	C6orf31	6p21.32	0.09921	0.08384	Y
1556029_s_at	NMNAT2	1q25	0.09921	0.08439	Y
221803_s_at	NRBF2	10q21.3	0.09621	0.09144	Y
202824_s_at	TCEB1	8q21.11	0.08788	0.0479	N
201848_s_at	BNIP3	10q26.3	0.08326	0.0479	N
203621_at	NDUFB5	3q26.33	0.08326	0.0479	N
223328_at	SVH	7q22.1	0.08326	0.0479	N
203830_at	NJMU-R1	17q11.2	0.08326	0.0479	N
223273_at	C14orf142	14q32.13	0.08326	0.0479	N
209096_at	UBE2V2	8q11.21	0.08439	0.0479	N
209036_s_at	MDH2	7p12.3-q11.2	0.08326	0.0479	N
202507_s_at	SNAP25	20p12-p11.2	0.08326	0.0479	N
218027_at	MRPL15	8q11.2-q13	0.08326	0.0479	N
205361_s_at	PFDN4	20q13.2	0.08326	0.0479	N
1567458_s_at	RAC1	7p22	0.08326	0.0479	N
224719_s_at	GRCC10	12p13.31	0.08326	0.0479	N
226338_at	DKFZp762O076	8q21.3	0.08326	0.0479	N
235916_at	YPEL4	11q12.1	0.08326	0.0479	N
228283_at	MGC61571	3p24.1	0.08326	0.0479	N
208868_s_at	GABARAPL1	12p13.2	0.08326	0.0479	N
209080_x_at	TXNL2	6p25.3	0.08326	0.0479	N
228614_at	LOC205251	2q13	0.09621	0.0479	N
202854_at	HPRT1	Xq26.1	0.08326	0.0479	N
207508_at	ATP5G3	2q31.1	0.09023	0.0479	N

The top 50 upregulated and top 50 downregulated transcripts are shown. For the complete list of transcripts, see [ ].

Supplemental Table 4. Genes differentially expressed in non-demented versus demented common to all three data sets.[a]

Upregulated in AD

Gene Symbol	Gene Name	q-value	chromosome location
MYO1C	myosin IC	0.004	17p13
NFKBIA	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha	0.005	14q13
GEM	GTP binding protein overexpressed in skeletal muscle	0.008	8q13-q21
SRRM2	serine/arginine repetitive matrix 2	0.011	16p13.3
RoXaN	rotavirus X protein associated with NSP3	0.011	22q13.2
RBM6	RNA binding motif protein 6	0.013	3p21.3
BACH2	BTB and CNC homology 1, basic leucine zipper transcription factor 2	0.014	6q15
ABCA1	ATP-binding cassette, sub-family A (ABC1), member 1	0.014	9q31.1
ITGB5	integrin, beta 5	0.016	3q21.2
PER1	period homolog 1 (Drosophila)	0.018	17p13.1-17p12
ITSN1	intersectin 1 (SH3 domain protein)	0.018	21q22.1-q22.2
TBX6	T-box 6	0.024	16p11.2
BCL2	B-cell CLL/lymphoma 2	0.033	18q21.33, 18q21.3
ALDH1A2	aldehyde dehydrogenase 1 family, member A2	0.037	15q21.3
PABPN1	poly(A) binding protein, nuclear 1	0.040	14q11.2-q13
IQGAP1	IQ motif containing GTPase activating protein 1	0.040	15q26.1
FOXO1A	forkhead box O1A (rhabdomyosarcoma)	0.043	13q14.1
ROCK1	Rho-associated, coiled-coil containing protein kinase 1	0.043	18q11.1
GRB10	growth factor receptor-bound protein 10	0.043	7p12-p11.2
HDAC7A	histone deacetylase 7A	0.051	12q13.1
CSNK1A1	casein kinase 1, alpha 1	0.052	5q32
CTBP2	C-terminal binding protein 2	0.052	10q26.13
ATP8B1	ATPase, Class I, type 8B, member 1	0.053	18q21-q22, 18q21.31
NCOA3	nuclear receptor coactivator 3	0.054	20q12
MACF1	microtubule-actin crosslinking factor 1	0.057	1p32-p31
GTSE1	G-2 and S-phase expressed 1	0.059	22q13.2-q13.3
PGCP	plasma glutamate carboxypeptidase	0.059	8q22.2
GPR107	G protein-coupled receptor 107	0.059	9q34.11
LAMA4	laminin, alpha 4	0.062	6q21
MAP2K7	mitogen-activated protein kinase kinase 7	0.065	19p13.3-p13.2
DDR2	discoidin domain receptor family, member 2	0.067	1q12-q23
ATBF1	AT-binding transcription factor 1	0.070	16q22.3-q23.1
SLC4A4	solute carrier family 4, sodium bicarbonate cotransporter, member 4	0.070	4q21
TIMELESS	Timeless homolog (Drosophila)	0.072	12q12-q13
FBLN1	fibulin 1	0.072	22q13.31
MXI1	MAX interactor 1 /// MAX interactor 1	0.073	10q24-q25
LAMC1	laminin, gamma 1 (formerly LAMB2)	0.075	1q31
FBXL11	F-box and leucine-rich repeat protein 11	0.075	11q13.2
LOC51035	ORF	0.075	11q12.3
BAZ2B	bromodomain adjacent to zinc finger domain, 2B	0.077	2q23-q24
TBL1X	transducin (beta)-like 1X-linked	0.082	Xp22.3
MAP2K3	Mitogen-activated protein kinase kinase 3 /// Mitogen-activated protein kinase kinase 3	0.083	17q11.2
HGF	hepatocyte growth factor (hepapoietin A; scatter factor)	0.084	7q21.1
PLCE1	phospholipase C, epsilon 1	0.090	10q23
USP22	ubiquitin specific protease 22	0.094	17p11.2
CDC14A	CDC14 cell division cycle 14 homolog A (S. cerevisiae)	0.095	1p21

LRRFIP2	leucine rich repeat (in FLII) interacting protein 2	0.098	3p22.3
YES1	v-yes-1 Yamaguchi sarcoma viral oncogene homolog 1	0.098	18p11.31-p11.21
TM4SF6	transmembrane 4 superfamily member 6	0.099	Xq22

Downregulated in AD

Gene Symbol	Gene Name	q-value	Chr. Location
ATP6V1G2	ATPase, H+ transporting, lysosomal 13kDa, V1 subunit G isoform 2	0.002	6p21.3
C20orf13	chromosome 20 open reading frame 13	0.004	20p12.1
PCMT1	protein-L-isoaspartate (D-aspartate) O-methyltransferase	0.004	6q24-q25
AUH	AU RNA binding protein/enoyl-Coenzyme A hydratase	0.004	9q22.31
HBXIP	hepatitis B virus x interacting protein	0.004	1p13.3
SGNE1	secretory granule, neuroendocrine protein 1 (7B2 protein)	0.005	15q13-q14
SLC25A6	solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 6	0.005	Xp22.32 and Yp
NDUFA5	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 5, 13kDa	0.005	7q32
SRD5A1	steroid-5-alpha-reductase, alpha polypeptide 1 (3-oxo-5 alpha-steroid delta 4-dehydrogenase alpha 1)	0.005	5p15
ACTR1A	ARP1 actin-related protein 1 homolog A, centractin alpha (yeast)	0.006	10q24.32
SYT1	synaptotagmin I	0.006	12cen-q21
LGALS8	lectin, galactoside-binding, soluble, 8 (galectin 8)	0.006	1q42-q43
PPP1R7	protein phosphatase 1, regulatory subunit 7	0.006	2q37.3
RGS7	regulator of G-protein signalling 7	0.006	1q43, 1q23.1
CALM3	calmodulin 3 (phosphorylase kinase, delta)	0.007	19q13.2-q13.3
RAB2	RAB2, member RAS oncogene family	0.007	8q12.1
CAMK1G	calcium/calmodulin-dependent protein kinase IG	0.007	1q32-q41
ACTR3	ARP3 actin-related protein 3 homolog (yeast)	0.007	2q14.1
PPP2R2B	protein phosphatase 2 (formerly 2A), regulatory subunit B (PR 52), beta isoform	0.008	5q31-5q32
ATP6V0C	ATPase, H+ transporting, lysosomal 16kDa, V0 subunit c	0.008	16p13.3
PSMB4	proteasome (prosome, macropain) subunit, beta type, 4	0.008	1q21
GAP43	growth associated protein 43	0.008	3q13.1-q13.2
PSMB2	proteasome (prosome, macropain) subunit, beta type, 2 /// proteasome (prosome, macropain) subunit, beta type, 2	0.008	1p34.2
CXX1	CAAX box 1	0.008	Xq26
ANXA7	annexin A7	0.008	10q21.1-q21.2
RAD51C	RAD51 homolog C (S. cerevisiae)	0.008	17q22-q23
SMARCA2	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 2	0.009	9p22.3
BEX1	brain expressed, X-linked 1	0.009	Xq21-q23
MEF2C	MADS box transcription enhancer factor 2, polypeptide C (myocyte enhancer factor 2C)	0.009	5q14
MAPRE3	microtubule-associated protein, RP/EB family, member 3	0.009	2p23.3-p23.1
SNX3	sorting nexin 3	0.010	6q21
SRPK2	SFRS protein kinase 2	0.010	7q22-q31.1
POLR2K	polymerase (RNA) II (DNA directed) polypeptide K, 7.0kDa	0.010	8q22.2
KCNJ3	potassium inwardly-rectifying channel, subfamily J, member 3	0.010	2q24.1
SYNJ1	synaptojanin 1	0.010	21q22.2
RIT2	Ras-like without CAAX 2	0.010	18q12.3
ARPP-19	cyclic AMP phosphoprotein, 19 kD	0.011	15q21.2
MRP63	mitochondrial ribosomal protein 63	0.011	
TIMM17A	translocase of inner mitochondrial membrane 17 homolog A (yeast)	0.011	1q32.1

RY1	putative nucleic acid binding protein RY-1	0.012	2p13.3
ATP2A2	ATPase, Ca++ transporting, cardiac muscle, slow twitch 2	0.012	12q23-q24.1
TPI1	triosephosphate isomerase 1	0.012	12p13
HSPA9B	heat shock 70kDa protein 9B (mortalin-2)	0.012	5q31.1
ZNF207	zinc finger protein 207	0.012	17q11.2
EIF2S1	eukaryotic translation initiation factor 2, subunit 1 alpha, 35kDa	0.012	14q23.3
LDOC1	leucine zipper, down-regulated in cancer 1	0.013	Xq27
ATP5C1	ATP synthase, H+ transporting, mitochondrial F1 complex, gamma polypeptide 1	0.013	10p15.1
BACH	brain acyl-CoA hydrolase	0.013	1p36.31-p36.11
MOCS2	molybdenum cofactor synthesis 2	0.013	5q11
ARF3	ADP-ribosylation factor 3 /// ADP-ribosylation factor 3	0.013	12q13
TDE1	tumor differentially expressed 1 /// tumor differentially expressed 1	0.013	20q13.1-13.3
CRI1	CREBBP/EP300 inhibitor 1	0.013	15q21.1-q21.2
PSME3	proteasome (prosome, macropain) activator subunit 3 (PA28 gamma; Ki)	0.013	17q21
PPP2R5C	protein phosphatase 2, regulatory subunit B (B56), gamma isoform	0.014	14q32
SNX10	sorting nexin 10	0.014	7p15.2
PSMC2	proteasome (prosome, macropain) 26S subunit, ATPase, 2	0.014	7q22.1-q22.3
GNAS	GNAS complex locus	0.015	20q13.2-q13.3
MTMR1	myotubularin related protein 1	0.015	Xq28
IARS	isoleucine-tRNA synthetase	0.015	9q21
RGS4	regulator of G-protein signalling 4	0.015	1q23.3
PTPRN2	protein tyrosine phosphatase, receptor type, N polypeptide 2	0.015	7q36
VDAC3	voltage-dependent anion channel 3	0.016	8p11.2
MATR3	matrin 3	0.016	5q31.2
SH3GLB2	SH3-domain GRB2-like endophilin B2	0.016	9q34
GPI	glucose phosphate isomerase	0.016	19q13.1
SFPQ	splicing factor proline/glutamine rich (polypyrimidine tract binding protein associated)	0.017	1p34.3
SUPT4H1	suppressor of Ty 4 homolog 1 (S. cerevisiae)	0.017	17q21-q23
PREP	prolyl endopeptidase	0.017	6q22
PAM	peptidylglycine alpha-amidating monooxygenase	0.018	5q14-q21
ATP5L	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit g	0.018	11q23.3
TMEFF1	transmembrane protein with EGF-like and two follistatin-like domains 1	0.018	9q31
CAPNS1	calpain, small subunit 1 /// calpain, small subunit 1	0.018	19q13.12
RAP1GDS1	RAP1, GTP-GDP dissociation stimulator 1	0.018	4q23-q25
ELMO1	engulfment and cell motility 1 (ced-12 homolog, C. elegans)	0.018	7p14.1
BSCL2	Bernardinelli-Seip congenital lipodystrophy 2 (seipin)	0.018	11q12-q13.5
NELL1	NEL-like 1 (chicken)	0.018	11p15.2-p15.1
ATP8A2	ATPase, aminophospholipid transporter-like, Class I, type 8A, member 2	0.018	13q12-13
SRPR	signal recognition particle receptor ('docking protein')	0.018	11q24.3
PIGK	phosphatidylinositol glycan, class K	0.019	1p31.1
FKBP1B	FK506 binding protein 1B, 12.6 kDa	0.019	2p23.3
RPL5	ribosomal protein L5	0.019	1p22.1
SCG3	secretogranin III	0.019	15q21
PLCB1	phospholipase C, beta 1 (phosphoinositide-specific)	0.019	20p12
GC20	translation factor sui1 homolog	0.020	3p22.1
VAMP1	vesicle-associated membrane protein 1 (synaptobrevin 1)	0.020	12p
YWHAZ	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide	0.021	8q23.1
BNIP3	BCL2/adenovirus E1B 19kDa interacting protein 3	0.021	10q26.3

SSB	Sjogren syndrome antigen B (autoantigen La)	0.021	2q31.1
VARS2	valyl-tRNA synthetase 2	0.022	6p21.3
CKMT1	creatine kinase, mitochondrial 1 (ubiquitous)	0.022	15q15
CCNA1	cyclin A1	0.022	13q12.3-q13
GNAO1	guanine nucleotide binding protein (G protein), alpha activating activity polypeptide O	0.022	16q13
UBE2N	ubiquitin-conjugating enzyme E2N (UBC13 homolog, yeast)	0.023	12q22
B4GALT6	UDP-Gal:betaGlcNAc beta 1,4- galactosyltransferase, polypeptide 6	0.023	18q11
RAB7	RAB7, member RAS oncogene family	0.023	3q21.3
MDH2	malate dehydrogenase 2, NAD (mitochondrial)	0.024	7p12.3-q11.2
CSE1L	CSE1 chromosome segregation 1-like (yeast)	0.024	20q13
TAF9	TAF9 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 32kDa	0.024	5q11.2-q13.1
PPP3CA	protein phosphatase 3 (formerly 2B), catalytic subunit, alpha isoform (calcineurin A alpha)	0.024	4q21-q24
UBC	ubiquitin C	0.024	12q24.3
ASAH1	N-acylsphingosine amidohydrolase (acid ceramidase) 1	0.025	8p22-p21.3
APP	amyloid beta (A4) precursor protein (protease nexin-II, Alzheimer disease)	0.025	21q21.2, 21q21.3
PSMB3	proteasome (prosome, macropain) subunit, beta type, 3	0.025	17q12
CUL2	cullin 2	0.025	10p11.21
RPL15	ribosomal protein L15	0.026	3p24.2
GPAA1	GPAA1P anchor attachment protein 1 homolog (yeast)	0.026	8q24.3
PEX7	peroxisomal biogenesis factor 7	0.027	6q21-q22.2
NDUFA4	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 4, 9kDa	0.027	7p21.3
QP-C	low molecular mass ubiquinone-binding protein (9.5kD)	0.027	5q31.1
FBXL2	F-box and leucine-rich repeat protein 2	0.027	3p23
RECQL	RecQ protein-like (DNA helicase Q1-like)	0.027	12p12
KIAA0436	putative prolyl oligopeptidase	0.027	2p22.1
RANBP9	RAN binding protein 9	0.027	6p23
NRXN1	neurexin 1	0.028	2p16.3
IDS	iduronate 2-sulfatase (Hunter syndrome)	0.029	Xq28
GRIA3	glutamate receptor, ionotrophic, AMPA 3	0.029	Xq25-q26
SYNGR1	synaptogyrin 1	0.029	22q13.1
ATP2B2	ATPase, Ca++ transporting, plasma membrane 2	0.029	3p25.3
ATP5O	ATP synthase, H+ transporting, mitochondrial F1 complex, O subunit (oligomycin sensitivity conferring protein)	0.030	21q22.1-q22.2, 21q22.11
OPA1	optic atrophy 1 (autosomal dominant)	0.030	3q28-q29, 3q28-q29
UBE2B	ubiquitin-conjugating enzyme E2B (RAD6 homolog) /// ubiquitin-conjugating enzyme E2B (RAD6 homolog)	0.030	5q23-q31
SCN2A2	sodium channel, voltage-gated, type II, alpha 2	0.030	2q23-q24
SH3BGRL	SH3 domain binding glutamic acid-rich protein like	0.031	Xq13.3
KPNB1	karyopherin (importin) beta 1	0.031	17q21.32
KATNB1	katanin p80 (WD repeat containing) subunit B 1	0.032	16q13
ESD	esterase D/formylglutathione hydrolase	0.033	13q14.1-q14.2
GABRB3	gamma-aminobutyric acid (GABA) A receptor, beta 3	0.033	15q11.2-q12
HSD17B12	hydroxysteroid (17-beta) dehydrogenase 12	0.034	11p11.2
SYN2	synapsin II	0.034	3p25
TOMM70A	translocase of outer mitochondrial membrane 70 homolog A (yeast)	0.034	3q12.2
OLFM1	olfactomedin 1	0.036	9q34.3
COX7A2	cytochrome c oxidase subunit VIIa polypeptide 2 (liver)	0.036	6q12
STAT1	signal transducer and activator of transcription 1, 91kDa	0.037	2q32.2
PINK1	PTEN induced putative kinase 1	0.037	1p36
H2AFY	H2A histone family, member Y	0.037	5q31.3-q32
RTN4	reticulon 4	0.037	2p16.3

ORC5L	origin recognition complex, subunit 5-like (yeast)	0.038	7q22.1
OSBPL10	oxysterol binding protein-like 10	0.038	3p22.3
SRP72	signal recognition particle 72kDa	0.040	4q11
HMGN4	high mobility group nucleosomal binding domain 4	0.041	6p21.3
MTHFD1	methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 1, methenyltetrahydrofolate cyclohydrolase, formyltetrahydrofolate synthetase	0.042	14q24
PPP3CB	protein phosphatase 3 (formerly 2B), catalytic subunit, beta isoform (calcineurin A beta)	0.042	10q21-q22
PSMD1	proteasome (prosome, macropain) 26S subunit, non-ATPase, 1	0.042	2q37.1
PPP1R11	protein phosphatase 1, regulatory (inhibitor) subunit 11	0.044	6p21.3
SNCG	synuclein, gamma (breast cancer-specific protein 1)	0.045	10q23.2-q23.3
DNAJC8	DnaJ (Hsp40) homolog, subfamily C, member 8	0.045	1p35.3
CBARA1	calcium binding atopy-related autoantigen 1	0.045	10q22.1
YME1L1	YME1-like 1 (S. cerevisiae)	0.045	10p14
PPP3R1	protein phosphatase 3 (formerly 2B), regulatory subunit B, 19kDa, alpha isoform (calcineurin B, type I)	0.047	2p15
STXBP1	syntaxin binding protein 1	0.047	9q34.1
STOML1	stomatin (EPB72)-like 1	0.047	15q24-q25
BBP	beta-amyloid binding protein precursor	0.049	1p31.3
NDUFB4	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 4, 15kDa	0.049	3q13.33
SYT5	synaptotagmin V	0.049	19q, 11p
HSPA8	heat shock 70kDa protein 8	0.049	11q24.1
PMPCB	peptidase (mitochondrial processing) beta	0.050	7q22-q32
GNB1	guanine nucleotide binding protein (G protein), beta polypeptide 1	0.050	1p36.33
TUBB	tubulin, beta polypeptide	0.050	6p21.33
PIK4CA	phosphatidylinositol 4-kinase, catalytic, alpha polypeptide	0.050	22q11.21
UBE2G1	ubiquitin-conjugating enzyme E2G 1 (UBC7 homolog, C. elegans)	0.050	1q42, 17p13.2
LXN	latexin	0.052	3q25.32
ARL7	ADP-ribosylation factor-like 7	0.053	2q37.1
VAMP2	vesicle-associated membrane protein 2 (synaptobrevin 2)	0.054	17p13.1
TUBB2	tubulin, beta, 2	0.054	6p25
RTN1	reticulon 1	0.054	14q23.1
RAB6A	RAB6A, member RAS oncogene family	0.055	11q13.3
PFDN4	prefoldin 4	0.056	20q13.2
MYT1L	myelin transcription factor 1-like	0.057	2p25.3
SYNGR3	synaptogyrin 3	0.058	16p13
PAI-RBP1	PAI-1 mRNA-binding protein	0.058	1p31-p22
STMN2	stathmin-like 2	0.058	8q21.13
NEF3	neurofilament 3 (150kDa medium)	0.059	8p21
PDE4DIP	phosphodiesterase 4D interacting protein (myomegalin)	0.060	1q12
SCAMP1	secretory carrier membrane protein 1	0.061	5q13.3-q14.1
TRIM32	tripartite motif-containing 32	0.062	9q33.1
VSNL1	visinin-like 1	0.062	2p24.3
TCTE1L	t-complex-associated-testis-expressed 1-like	0.063	Xp21
GNAL	guanine nucleotide binding protein (G protein), alpha activating activity polypeptide, olfactory type	0.063	18p11.22-p11.21
IDH3B	isocitrate dehydrogenase 3 (NAD+) beta	0.065	20p13
GAD2	glutamate decarboxylase 2 (pancreatic islets and brain, 65kDa)	0.065	10p11.23
STK24	serine/threonine kinase 24 (STE20 homolog, yeast)	0.066	13q31.2-q32.3
TTC1	tetratricopeptide repeat domain 1	0.069	5q32-q33.2
COX11	COX11 homolog, cytochrome c oxidase assembly protein (yeast) /// COX11 homolog, cytochrome c oxidase assembly protein (yeast)	0.069	17q22
ARPC2	actin related protein 2/3 complex, subunit 2, 34kDa	0.071	2q36.1

TA-LRRP	T-cell activation leucine repeat-rich protein	0.072	1p22.2
CACNB2	calcium channel, voltage-dependent, beta 2 subunit	0.074	10p12
TBCA	tubulin-specific chaperone a	0.074	5q14.1
CDC27	cell division cycle 27	0.078	17q12-17q23.2
SNRPA1	small nuclear ribonucleoprotein polypeptide A'	0.078	15q26.3
R3HDM	R3H domain (binds single-stranded nucleic acids) containing	0.082	2q21.3
ELAVL4	ELAV (embryonic lethal, abnormal vision, Drosophila)-like 4 (Hu antigen D)	0.088	1p34
DUSP6	dual specificity phosphatase 6	0.089	12q22-q23
ALDOC	aldolase C, fructose-bisphosphate	0.091	17cen-q12
TMEM4	transmembrane protein 4	0.095	12q15
SLC25A12	solute carrier family 25 (mitochondrial carrier, Aralar), member 12	0.095	2q24
RSU1	Ras suppressor protein 1	0.097	10p13
CDC5L	CDC5 cell division cycle 5-like (S. pombe)	0.098	6p21
ATP6V1F	ATPase, H+ transporting, lysosomal 14kDa, V1 subunit F	0.099	7q32
RAB11A	RAB11A, member RAS oncogene family	0.100	15q21.3-q22.31

[a] Differentially expressed transcripts are $q < 0.1$ from ANOVA

Supplemental Table 5. Transcripts differentially expressed in NFT formation that are also found in Dunckley et al.[a]

Upregulated in high Braak

Probe Set ID	Symbol	Gene	q-value	p-value Dunckley, et al
232286_at	LAF4	Lymphoid nuclear protein related to AF4	0.0833	0.0335
201502_s_at	NFKBIA	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha	0.0879	0.0731
235305_s_at	FLJ10948	Hypothetical protein FLJ10948	0.0907	0.0469
239892_at	RARS	Arginyl-tRNA synthetase	0.0833	0.0867
233626_at	NRP1	Neuropilin 1	0.0833	0.0796
222186_at	ZA20D3	Zinc finger, A20 domain containing 3	0.0833	0.0263
215109_at	KIAA0492	KIAA0492 protein	0.0833	0.0928
214656_x_at	MYO1C	myosin IC	0.0833	0.0757
1554963_at			0.0833	0.0926
228214_at			0.0947	0.0542

Downregulated in high Braak

Probe Set ID	Symbol	Gene	q-value	p-value Dunckley, et al
203415_at	PDCD6	programmed cell death 6	0.0833	0.0252
202854_at	HPRT1	hypoxanthine phosphoribosyltransferase 1 (Lesch-Nyhan syndrome)	0.0833	0.0091
202930_s_at	SUCLA2	succinate-CoA ligase, ADP-forming, beta subunit	0.0833	0.0407
204905_s_at	EEF1E1	eukaryotic translation elongation factor 1 epsilon 1	0.0833	0.0756
207235_s_at	GRM5	glutamate receptor, metabotropic 5	0.0833	0.0229
241998_at		Similar to RIKEN cDNA D630023F18	0.0833	0.0094
230839_at	HRMT1L4	HMT1 hnRNP methyltransferase-like 4 (S. cerevisiae)	0.0833	0.0418
221805_at	NEFL	neurofilament, light polypeptide 68kDa	0.0833	0.0022
228062_at	NAP1L5	nucleosome assembly protein 1-like 5	0.0833	0.0144
203594_at	RTCD1	RNA terminal phosphate cyclase domain 1	0.0833	0.039
203830_at	NJMU-R1	protein kinase Njmu-R1	0.0833	0.0961
205123_s_at	TMEFF1	transmembrane protein with EGF-like and two follistatin-like domains 1	0.0833	0.0358
232305_at	HMGCLL1	3-hydroxymethyl-3-methylglutaryl-Coenzyme A lyase-like 1	0.0833	0.0946
236738_at	LOC401097	Similar to LOC166075	0.0833	0.0421
206339_at	CART	cocaine- and amphetamine-regulated transcript	0.0833	0.0735
209303_at	NDUFS4	NADH dehydrogenase (ubiquinone) Fe-S protein 4, 18kDa (NADH-coenzyme Q reductase)	0.0838	0.039
209096_at	UBE2V2	ubiquitin-conjugating enzyme E2 variant 2	0.0844	0.077
222230_s_at	ACTR10	actin-related protein 10 homolog (S. cerevisiae)	0.0879	0.0139
233135_at		Homo sapiens, clone IMAGE:5199801, mRNA	0.0879	0.0505
218545_at	FLJ11088	GGA binding partner	0.0891	0.0631
201823_s_at	RNF14	ring finger protein 14	0.0891	0.0753
229506_at		CDNA clone IMAGE:5263177, partial cds	0.0900	0.0259
223503_at	DKFZP566N034	Hypothetical protein DKFZp566N034	0.0956	0.0439
204807_at	TMEM5	transmembrane protein 5	0.0962	0.0359
227669_at	DKFZP564B167	DKFZP564B167 protein	0.0962	0.043
212434_at	GRPEL1	GrpE-like 1, mitochondrial (E. coli)	0.0974	0.0773
212551_at	CAP2	CAP, adenylate cyclase-associated protein, 2 (yeast)	0.0974	0.0224
213149_at	DLAT	dihydrolipoamide S-acetyltransferase (E2 component of pyruvate dehydrogenase complex)	0.0995	0.0069
231102_at	CROT	carnitine O-octanoyltransferase	0.0996	0.0274

[a] Differentially expressed transcripts are $q < 0.1$ from ANOVA

Chapter III. Impact of Genotype call Rate on Complex Trait Studies: Low-Level Analysis of Genotyping Algorithms on the Affymetrix 100K Gene Mapping set.

Introduction

Advances in the ability to identify Human DNA sequence variation across the genome have led to improved understanding of the relationship of genetic variation to human complex phenotypes and have highlighted the need for the proper tools to identify all types of DNA variation accurately and precisely with high sensitivity [2, 5, 7, 121]. Common genetic variants influence gene expression levels in individuals [122-124] and in some instances are involved in susceptibility to disease [3, 30, 121, 125]. In addition to variation in expression, variation at the sequence level may be due to polymorphisms in copy number variable regions. Genome wide techniques for studying genetic association to disease rely on the extent of linkage disequilibrium (LD) between the SNP marker tested and the genetic alteration involved in susceptibility [4-6]. The highly parallel nature of these technologies demands extremely accurate and complete genotype calling algorithms. It is important to assess the impact of the genotype call rate on complex trait studies. Therefore, we investigated the low-level performance of the Affymetrix GeneChip 100K mapping set algorithm [34] to determine the sensitivity of the algorithm to noise, chip features, experimental variation and sample characteristics. We also compared the performance of the standard Affymetrix Dynamic Modeling algorithm with the recently released BRLMM algorithm Affymetrix instituted for the genotyping the 500K mapping set (http://www.affymetrix.com/support/technical/whitepapers/brlmm_whitepaper.pdf, http://www.broad.mit.edu/gen_analysis/genotyping/brlmm_affy_ncrr.html)

2. Methods:

2A. DNA and Hybridization:

DNA from subjects with a clinical diagnosis of no dementia within a year of death (N= 30) or Probable Alzheimer's disease (N= 25) [65]was used in this study. All subjects were participants in studies performed by the Layton Aging and Alzheimer 's disease Center, Portland, Oregon and included the 14 subjects from the previous gene expression profiling study (chapter II) DNA was isolated from either whole blood or postmortem human frontal cortex brain tissue. DNA from whole blood was isolated using the QIAmp DNA blood kit (Qiagen, Valencia, CA). For deceased subjects, approximately 100 mg of brain tissue (previously frozen at -80^{o}C) was processed for genomic DNA using the Wizard Genomic DNA purification Kit (Promega, Madison WI) following manufacture's instructions.

Isolated genomic DNA from each subject was digested and labeled following manufacturer's instructions (Affymetrix Inc., Santa Clara, CA). Briefly, 250ng of genomic DNA was digested with a restriction enzyme (XbaI or HindIII), ligated to an appropriate adapter for each enzyme, and amplified by PCR using a single primer. The PCR products were then digested with DNaseI, labeled and hybridized separately to the Affymetrix GeneChip Mapping 100K array chips. The arrays were scanned and genotypes called by the DM and BRLMM algorithms. These arrays contain probe sets to interrogate 58960 (XbaI) and 58974 (HindIII) SNPs across the entire human genome. Sample labeling and array hybridizations and processing were performed in the Affymetrix Microarray Core, Gene Microarray Shared Resource, Oregon Health & Science University.

2B. Genotyping Algorithms

2.B1. Dynamic Modeling Algorithm (DM)

The sensitivity of the DM algorithm was assessed through simulation and with original data. For simulation studies, each step of the DM algorithm as described [34] was implemented in the R v2.3.1 system for statistical computation [126]. Packages included in the Comprehensive R Archive Network (CRAN, http://cran.r-project.org/) were utilized for specific analyses as described below.

Custom scripts were written to follow the steps of the DM algorithm and create summary tables for each simulation condition at each step of the algorithm. Data obtained from DNA hybridized SNP chips were used to determine the hybridization intensity distribution. All SNPs for each called genotype were used to fit the genotypes separately to distributions utilizing the R package MASS v4 function fitdistr[127]). The lognormal distribution was chosen for all simulations.

Variables tested in the simulations included: overall chip intensity, background intensity, feature pixel number, and feature pixel intensity standard deviation (SD) (Table 1). Custom scripts were written to simulate 10 probe quartets for each overall chip intensity (high, medium, low) and background intensity (high, low). All combinations of these data sets (six in total) were used separately to simulate different pixel numbers and different SD so that the only difference among data sets was the variable to be tested. A total of 100 SNPs were simulated for each condition.

The final step in the DM genotyping algorithm is determining the SNP call based on the lowest p-value from the Wilcoxon signed rank test. In addition, a custom summary score was created and compared to the DM summarization. The summary score of a probe quartet was the total count for each genotype model among the 10 SNP quartets for each SNP. The genotype model with the highest summary score was labeled the genotype call for that SNP [34].

In order to test the sensitivity of the DM algorithm to sample characteristics and chip features, the recommended settings of DM in the GType operating system (Affymetrix) were used to genotype DNA samples hybridized to both HindIII and XbaI chips.

2.B2. BRLMM:

The same DNA samples were also genotyped using the BRLMM algorithm as recommended for the Affymetrix GeneChip 500K mapping set (http://www.affymetrix.com/support/technical/whitepapers/brlmm_whitepaper.pdf) with the following modifications: XbaI and HindIII digested DNA samples were genotyped by the DM algorithm (Affymetrix GType software) at a sensitivity threshold of 0.25. Samples with DM call rates > 80% were combined into one data set per chip type and the BRLMM algorithm applied to each data set. All further analyses were performed on the DM genotyped and the BRLMM genotyped samples.

2.C. SNP Genotype Call characteristics:

The DM algorithm calls as implemented in R were compared to the simulated SNP genotype calls and the simulated SNP characteristics to determine the sensitivity of the DM algorithm to the various conditions. In order to assess the effectiveness of the BRLMM algorithm, the performance on the biological data set of the BRLMM algorithm was compared to the DM performance relative to chip features, experimental conditions and sample characteristics.

Characteristics of the SNPs per sample were tested for NoCalls (missing data) across samples and across SNPs using hierarchical cluster analysis (Hmisc v3.0 package for R, available on the CRAN and [128]) in which the fraction of NoCalls in common between any two SNPs was used as the similarity measure. Correlations among continuous variables were also investigated through hierarchical cluster analysis using [1-(correlation among all samples)] as the distance metric.

To determine if the distribution of NoCalls was dependent on neighboring SNPs, a Chi-square test of the number of NoCalls across all chromosomes was performed. The number of samples with NoCalls was dichotomized into two extreme groups of SNPs with <5 samples with NoCall and SNPs with >10 samples with NoCalls and the patterns across all chromosomes analyzed by Chi-square analysis.

SNPs were tested for deviation from Hardy-Weinberg equilibrium expectation (Genetics for R , [129]) using the exact test [129]. Samples were examined within the control and AD phenotypes and the false discovery rate (FDR, [130]) was calculated to correct for multiple testing.

2.D. PCR Fragment length

DNA fragment size information for every SNP was obtained from the Mapping 100K annotation file (http://www.affymetrix.com/support/technical/annotationfilesmain.affx). SNPs were binned by DNA fragment length and analyzed for a greater number of NoCalls across samples per SNP than would be expected by chance using chi-square analysis.

2.E. Concordance of underperforming SNPs with SNPs involved in copy number variation

Copy number estimation of genomic regions was performed using the Copy Number Analyser for GeneChip (CNAG v2) [131]. A region of genomic copy number change is determined relative to a chosen reference array set. Copy number regions are estimated for each array individually in comparison to non-demented samples (see chapter 3). Briefly, CNAG selects the maximum number of sample references such that the standard deviation of intensity values across all SNPs on chip is minimized. Therefore, from the data set of all non-demented subjects, each array for an AD subject is matched with different control arrays.

Pearson's Chi-Square test was used to examine the impact of SNPs with possible genotyping errors on copy number estimation. As a surrogate for genotyping errors, NoCall SNPs were used. SNPs not called by the DM and BRLMM algorithms were compared to regions of copy number change across the genome under the null hypothesis that there is no association between NoCall SNPs and CNV regions. The alternative hypothesis is that there are more differences in the numbers of NoCall SNPs located in CNV regions than would be expected by chance.

3. Results:

3.A. DM Model Simulations

A summary of all simulations shows that the performance of the DM algorithm is most affected by the overall chip intensity and the background intensity (Table 2). High overall chip intensity and a low background were the conditions under which the DM algorithm performs best (maximum concordance rate of 84%). Low overall intensity and high background had the highest rates of SNPs that could not be discriminated by the DM algorithm (maximum concordance rate of 15%). Both homozygous and heterozygous simulated SNPs were affected by the same conditions. Calls for heterozygous SNPs were more affected (maximum concordance rates of 72 vs. 84 for the homozygous SNPs).

Because the individual probe quartets within a SNP often showed the correct call even for low intensity chips (21-48%), a summary score of probe quartet genotype calls was created. The DM algorithm genotype calls were compared to a genotype call based on a summary score where the genotype model most often called among the 10 SNP quartets is labeled the genotype (Table 3). The DM call is based on p-values and so only SNPs with significant p-values are given a genotype call. Both homozygous and heterozygous simulated SNPs were called correctly more often by the summary score than the DM algorithm. Calls for SNPs simulated with high and medium intensity were increased to 85-100% call rates (17-66% improvement). Calls for low intensity chips were also improved 69 - 72% but the overall call rates were lower (78 - 87 %). High background affected the SNP calls to a greater extent across all chip intensities reducing the call rates for both DM and the summary score by 6-50%. Low intensity chips have the greatest impact on SNP call rates in which only 15% of the homozygous SNPs could be genotyped by the DM algorithm. Heterozygous SNPs could not be called correctly by the DM algorithm on low intensity chips. In cases where there were errors in calls, probe quartets within heterozygous simulated SNPs more often called a homozygous genotype than a null model (Table 3B). At medium chip intensity/high background and for low chip intensities, probe quartets called a homozygous model more often (49-78% vs. 12-48% for correctly called heterozygous).

Although the summary score increased the number of correctly called genotypes, it also increased the false positive call rate (Figure 1). The summary score calls genotypes irrespective of model p-value. A comparison of the SNP calls for the summary score versus p-values and conditioned by DM call shows that all DM genotyped calls are genotyped correctly by the summary score. DM NoCalls are most often called correctly by the summary score for both homozygous and heterozygous SNPs (92% and 88%, respectively). However, the summary score has a false positive rate of 21% across all simulations (Figure 1 and Table 3).

3.B. Comparison of DM and BRLMM genotyping algorithms

Performance based on restriction enzyme digest was examined in genotype calls for a sample of 55 XbaI and HindIII chips (DM threshold 0.25). XbaI chips had lower call rates overall (93.59%, CI(92.82, 94.36) vs. 97.73%, CI(97.17, 98.28) for HindIII and higher variability (SD=3.22 vs. 2.12) (Table 4, Figure 2).

XbaI and HindIII chips (DM call rate of ≥ 80%) were examined separately with the BRLMM algorithm. The call rates of both chip types were increased with the BRLMM algorithm (XbaI= 98.53%, 95% CI (98.27, 98.79) and HindIII = 99.26% , 95% CI (98.91, 99.61)) and the variability was reduced (SD =1.01 and 1.34 for XbaI and HindIII, respectively). XbaI hybridized chips have a lower intensity than HindIII chips (10.71, 95% CI (10.63, 10.79) vs. 10.96, 95% CI (10.89, 11.04)) (Table 4, Figure 2).

Although the BRLMM algorithm is more efficient at calling genotypes, the XbaI hybridizations are still more variable than the HindIII hybridizations. The XbaI chips also have lower overall hybridization intensity which highlights the relationship between low intensity and higher variance.

3.C. Effect of sample characteristics on call rates.

Sample characteristics were summarized by sample (Table 5). Overall chip intensity had the most effect on the call rate (Figure 3). Low chip intensity was related to lower call rates in both XbaI and HindIII samples. Variability in call rates and chip intensity was seen for chips hybridized on the same day (Figure 3a and 3b). Higher variability in XbaI chips call rates is seen within hybridization dates (Figure 3b).

Sample source and phenotype were examined and showed no effect on call rate (Figure 4). Overall chip intensity was correlated with the call rate while the PCR yield was more correlated with the discordance between the DM and BRLMM algorithms (Figure5).

3.C.1. Characteristics of the NoCalls across samples

Characteristics of the NoCall data were examined by number of SNPs with NoCalls per sample (Figure 6) and by correlation across samples (Figure 7). Both the DM and BRLMM algorithms showed the higher variability of the NoCalls among the XbaI samples (Figure 6), although the BRLMM algorithm consistently reduced the number of NoCalls. The HindIII chips had the lowest NoCall rate with the BRLMM algorithm (50 out of 55 chips were less than 1000 NoCalls).

The patterns of NoCall SNPs were examined by clustering across all samples using only NoCall data (Figure 7). The similarity measure for clustering was the fraction of NoCalls in common between any two samples. The HindIII samples that have the highest NoCall rate cluster together while the HindIII samples with lower NoCalls do not show a pattern of similar NoCall SNPs when examined across all samples when genotyped with the BRLMM algorithm. XbaI digested samples show a pattern of similarity based on both numbers of NoCalls and similarity of NoCalls per SNP across samples for both the DM and BRLMM algorithms (Figure 7). We tested whether the pattern of NoCalls was related to phenotype by conditioning on case/controls status in the DM genotyped samples (Figure 8). Similar NoCall patterns were seen for controls and AD cases in that samples were related to NoCalls by both number of NoCall SNPs and concordance of NoCall SNPs.

All calls in each sample were compared between the DM and BRLMM genotyping algorithms (Table 6). Discordant calls ranged between 526 and 11,986 and were correlated to the call rates; higher call rates led to higher concordant calls. Genotyping calls can be discordant due to either a difference in call between the two algorithms or a call and a NoCall. The difference in call rates between the two algorithms was generally due to the BRLMM algorithm's ability to call SNPs that the DM algorithm did not. This can be seen in the difference between total discordant calls and the proportion of discordant calls that are genotyped by both algorithms. For example, of 4426 discordant calls for sample 464, 4243 SNPs were not called by one algorithm and called by the other. Only 183 were genotyped by both algorithms. The proportion of SNPs that were discordant and called by both algorithms ranged from 0.024-0.067 with the lowest call rates having the highest non-NoCall discordant rate (0.067 discordance and 80% call rate).

Possible pairwise combinations of discordant calls between the two algorithms are:

1) call/ NoCall,

2) homozygous / heterozygous

3) homozygous AA / homozygous BB.

Of the discordant calls genotyped by both algorithms not NoCall, all were due to one algorithm calling a heterozygote and the other a homozygote (Table 6). For example, of the 183 discordant but not NoCall calls for sample 464, all of these are called heterozygote by one algorithm and homozygote by the other, resulting in 4.1% of the discordant SNPs (0.3% of the total number of SNPs on the Chip) for this sample discrepant for genotype.

The distributions of NoCalls across all samples per SNP were determined (Figure 9). Both chips showed most SNPs have only a few number of samples with NoCalls. However, there are SNPs that are not genotyped across a number of samples. It is also noted that XbaI digested samples show the most variability.

3.C2. Effect of chromosome position and DNA fragment length on NoCalls

Although most SNPs have only a few samples with NoCalls, some SNPs show a tendency for NoCalls across multiple samples and it is important to know whether the SNPs with many samples of NoCall are located near one another. Therefore, we investigated whether the neighboring SNPs had an effect on the NoCall rate.

Based on the XbaI chip and genotypes from the DM algorithm, there is a significant difference between SNPs with low numbers of NoCall samples ($\leq$5) and those with a higher number of samples ($\geq$10) not called (p-value = 0.001, Table 7). However, this is due to fewer NoCalls than expected for all but chromosomes 5 and 6. Chromosome 6 shows the highest number of NoCalls across all samples with the DM algorithm on XbaI chips. This chromosome does not show a difference with the BRLMM algorithm (data not shown). Chi-square analysis of XbaI DM NoCall SNPs binned across chromosome 6 showed no significant difference between low and high numbers of samples with NoCall SNPs (p-value = 0.3080).

We were also interested in the structure of the DNA fragments used to PCR amplify the fragments for hybridization. Length of DNA fragment has been shown to increase variability in some studies [131, 132, 133, 2005 #157]. We tested whether the number of SNPs with NoCalls across multiple samples was related to PCR fragment length by grouping the SNPs into bins depending on the size of the XbaI restriction enzyme fragment length on which they are located. The XbaI digested DNA fragments ranged for 198 to 2120 bp and were binned into groups of about 250 bp apart (Table 8). Chi-square analysis showed that as fragment length increased, the number of SNPs with NoCalls across $\geq$10 samples was greater than expected by chance. The BRLMM algorithm showed the same result, although many fewer SNPs were affected by the fragment length bias.

Because of the difficulty of the DM to distinguish between some of the heterozygote and homozygote calls, we next tested if the fragment length was related to genotype calls. Four individual samples of various call rates and chip intensities were tested by chi-square analysis (Table 9). For each sample, as the fragment increased there was an increase in the number of NoCalls observed and a corresponding decrease in the number of heterozygote calls above what would be expected by chance (Table 9A). The BRLMM algorithm was able to distinguish heterozygotes except for the largest fragment length for sample 1433B3 which had the lowest call rate (92.81%) and the lowest overall chip intensity (10.34).

3.C3. Characteristics of the SNP calls across all samples

The proportions of the calls determined by both algorithms are similar for the homozygous genotypes (Table 10). However, the DM algorithm has more difficulty calling heterozygotes. The BRLMM algorithm “rescues” the heterozygote calls as shown by the increase in AB call rate with the BRLMM algorithm. This results in a decrease in the number of SNPs that are invariant across all samples (Table 10).

One quality control measure for the whole genome association studies using high density SNPs is to remove SNPs that are out of Hardy-Weinberg equilibrium (HWE) under the assumption that there are genotyping errors in one of more samples [134, 135]. HWE was calculated for all samples as well as AD and controls separately (Table12). Control samples had a higher number of SNPs out of HWE (80-232) than AD cases (42-109) after correcting for multiple testing (Table 11A). The number of SNPs significantly out of HWE decreased 1.2 - 4.3% when SNPs for which >1 or 2 samples (Table 11B) had a NoCall were removed from the analysis.

Because of the relationship between the NoCalls and the fragment length, we tested whether the SNPs significantly out of HWE were more likely to be those on longer fragment lengths. A chi-square analysis showed that an excess of SNPs significantly out of HWE (p-value <0.05) were observed on larger fragment lengths (Table 12). This was seen in all SNPs out of HWE and uncorrected for multiple testing and also in controls (Table 12A and B). AD cases did not show this bias (Table 12C). Fragment length and intensity values were significantly associated for four individual samples of various overall chip intensities (10.34 - 11.99) and DM call rates (82.92 – 97.39) (Table 13). In all four samples, intensity values per SNP decreased as the fragment length increased.

Smaller fragments (<1160 bp) tended to have higher than expected intensity values (>2975) and larger fragments (>1400) tended toward the lower intensities (<2417).

The relationship between NoCalls and SNPs significantly out of HWE was investigated by chi square analysis (Table 14).   SNPs significantly out of HWE ($p < 0.05$) were more likely to be SNPs with higher numbers of NoCalls across all samples (Table 14A). SNPs with <5 NoCalls across all samples were less likely to be out of HWE than expected by chance where as SNPs with 5-35 NoCalls across all samples were more likely to be out of HWE than expected by chance   Unlike the fragment length, the association of HWE with NoCalls was true for both the controls and the AD cases (Table 14B and C).

3C4. Underperforming SNPs impact downstream analyses.

A total of 720 SNPs on the Affymetrix GeneChip Mapping 100K set were located within CNV regions of which, there were 219 DM NoCall SNPs and 74 BRLMM NoCall SNPs.   NoCall SNPs from both genotyping algorithms were significantly more likely to be located within CNV regions than expected by chance (DM p-value = 3.706 x10^{-12} and BRLMM p-value = 2.029x 10^{-14}) indicating a potential relationship between CNV and genotype call rate.

4. Discussion:

High throughput genotyping technologies promise to dramatically increase our ability to detect associations between phenotypic traits and genetic variants. However, the data generated by these methods must be examined closely for inaccuracies as even a small percentage has a large effect on the total data set. Genotype calling seems a deceptively simple task. In reality, this is a complex process involving a number of assumptions and summarizations over multiple probe sets per SNP. The DM algorithm is the standard genotyping algorithm for the Affymetrix 100K Mapping Set. Many association studies use this platform and are therefore subject to the inherent difficulty with calling heterozygote SNPs. Therefore, we set out to determine:

1) the sensitivity of the DM to chip features in order to discover in which instances the DM algorithm has difficulty making genotype calls and

2) whether the BRLMM algorithm recommended for the 500K Mapping Set can be used to improve the genotyping call on the 100K Mapping GeneChips.

Sensitivity of the DM algorithm to chip features

Simulation studies demonstrated the DM algorithm's dependence on the overall chip intensity and the relationship between intensity and variability. The DM algorithm shows best performance when overall chip intensity is high and the background is low. Variability is greater when the chip intensity is low. Heterozygous SNPs are more problematic for the DM algorithm than the homozygous SNPs and are more often not called (higher probability of the null model or low confidence in p-value for a genotype model). Other chip features such as pixel number and standard deviation do not affect the genotype call rate.

The DM algorithm uses a non-parametric summarization method to determine the final SNP call. Thresholds for the Wilcoxon signed rank test p-values change the stringency of the call rate. At an unadjusted p-value <0.05, no false positive calls were seen. A summary score summarization method across all genotype models improves the genotype calling rate but at the cost of increasing the false positive call rate. The main difficulty for the DM algorithm lies with the discrimination of the heterozygote SNPs which is directly related to chip intensity through the log likelihood model assumptions. This becomes increasingly problematic as the genotyping chips become denser because a small percentage of NoCalls translates into an ever increasing number of missing SNP calls.

BRLMM algorithm improves the genotyping call rate

Because the BRLMM algorithm is substantially different from the DM algorithm and does not rely on the same log likelihood models, we compared the sensitivity of the DM and BRLMM algorithms. Experimental data on the performance of the two Affymetrix chip types showed that the main difference in call rates is between the XbaI and HindIII chips and is due to the heterozygote dropout seen with the DM genotyping algorithm. Sample characteristics had no affect on genotyping calls for either algorithm.

Overall chip intensity and PCR fragment length had the most important effects on chip call rates. Chip intensity is related to the effeciency of the restriction enzyme digestion [26] and the PCR fragment amplification [131, 132] of the genomic DNA. Hybridization efficiency to the probes is related to the GC content of each specific probe and the GC content of the digested fragment [131, 132]. Higher GC content probes do not hybridize well to the target. Higher GC content target fragments and higher fragment length affect PCR efficiency. In our experiments, the longer PCR fragment lengths were related to higher NoCall rates. Lower overall chip intensity occurs when higher levels of the longer fragment lengths are not amplified in the PCR. SNPs located on longer PCR fragment lengths tend to be out of HWE more than expected by chance because they are more often not called due to DM algorithm’s difficulty in distinguishing genotypes with the lower intensity levels of these SNPs.

The BRLMM algorithm partially corrects for these probe specific effects in the first steps of the algorithm where quantile normalization across all arrays and allele specific intensities are modeled to summarize probe intensities from a probe set into a single value ([31], http://www.affymetrix.com/support/technical/whitepapers/brlmm_whitepaper.pdf). This is reflected in the reduced variance, improved call rate and fewer numbers of SNPs out of HWE.

The influence of NoCall SNPs on CNV regions was used as a surrogate for estimating the effect of genotyping errors on downstream analyses. Significant differences in the proportion of CNV variants between those with NoCall SNPs and those without highlight the difficulty of genotyping SNPs in regions of genomic complexity. This difference may be for several reasons. The greater number of NoCall genotypes in these CNV regions may reflect the complexity of the underlying chromosomal structure. It may be harder to type the SNPs located in CNV regions due to possible variability in SNP genotype for the different copies [136]. The more sophisticated BRLMM algorithm is able to call more SNPs with more subtle variations in intensity values among genotypes. Thus, the genotypes in these regions would be represented by the genotype of the most numerous alleles and not a reflection of the complete allele pattern in the multiple copies.

Accurate genotyping is critical for the ability to answer interesting biological questions about the effect of genetic sequence variation on phenotypes. Low-level analysis helps identify the strengths and weaknesses and overall performance of a genotyping algorithm. We have shown that SNPs with NoCalls are generally found across the genome and across samples at a low percentage and that the algorithms are sensitive to specific experimental vagaries and not to particular sample characteristics. This implies that samples with overall low call rates are most likely different in other ways. They may be possibly contaminated or they may have more interesting genetic variability such as copy number variations. CNV regions do not meet the assumption of diploidy for the genotyping algorithms and would be called as NoCalls. Other DNA sequences that do not hybridize well to the probes such as stretches of G/C or A/T or SNPs elsewhere in the target sequence may increase the NoCall rate as well as the cause false positive copy number estimations. The practice of discarding samples with low call rates upon repeated hybridization should be reconsidered as it may be abandoning samples with interesting genetic phenomena.

We have shown that the BRLMM algorithm improves the call rates for the 100K mapping array set and that overall chip intensity due to the variation in PCR fragment length is the most important experimental variable that affects call rates. This variability could be due to the PCR amplification step of the PCR product fractionation step. Several more issues need to be addressed to improve performance. The effect of combining chips into the BRLMM analysis from different hybridization batches was not investigated. This may introduce variation into the genotyping clusters and reduce genotyping call rates. In addition, more work needs to be done to investigate probe specific effects on call rates and the consequence of removing problematic probes from the analyses. As the technology advances and arrays become progressively more dense, it will be important to continually re-evaluate the sources of variability for each different platform.

These results provide a basis for studying the impact of sequence variation on genetic association studies of SNPs and copy number variations to susceptibility to disease as well as the effects of CNV on gene expression (Chapter IV).

Chapter III Figures and Tables.

Figures.

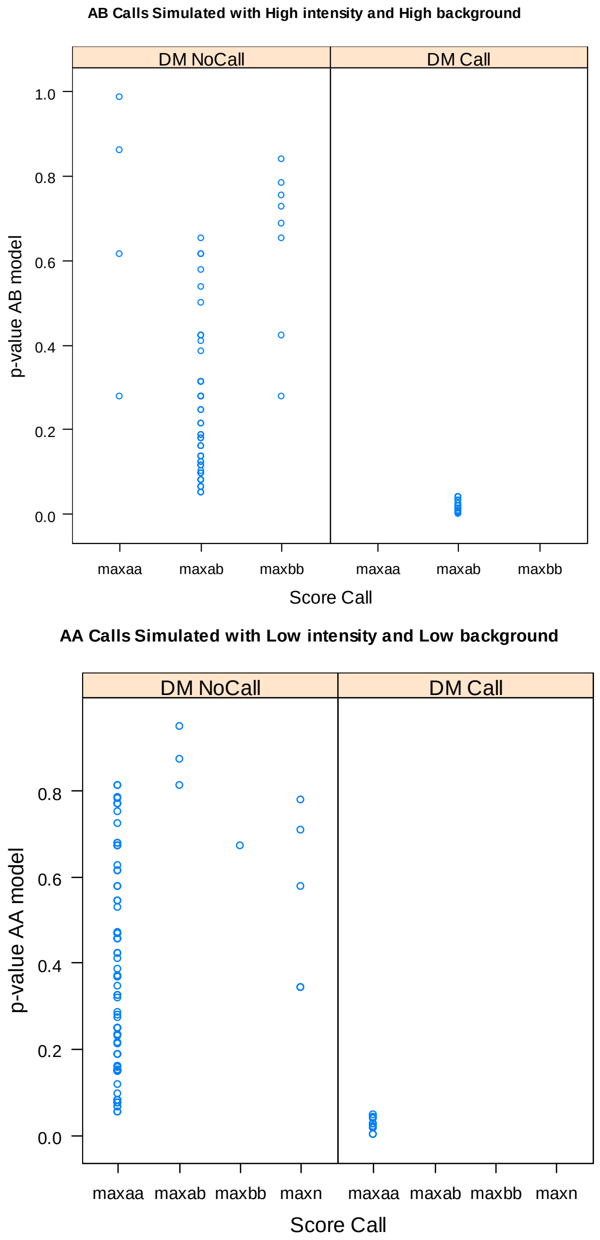


Figure 1. Score Calls for SNPs relative to p-value and DM call. For each simulated SNP, the maximum number of probe quartets called for each genotype model is given (maxaa, maxab, maxbb, maxn) relative to whether the SNP was called by the DM algorithm (DM call vs DM NoCall).

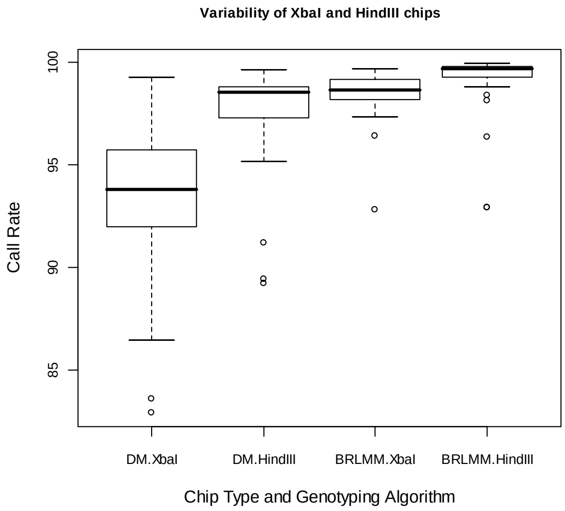


Figure 2. Variability of XbaI and HindIII chips genotyped by the DM or the BRLMM algorithms. Genotypes were called for all samples with DM call rate >80%.

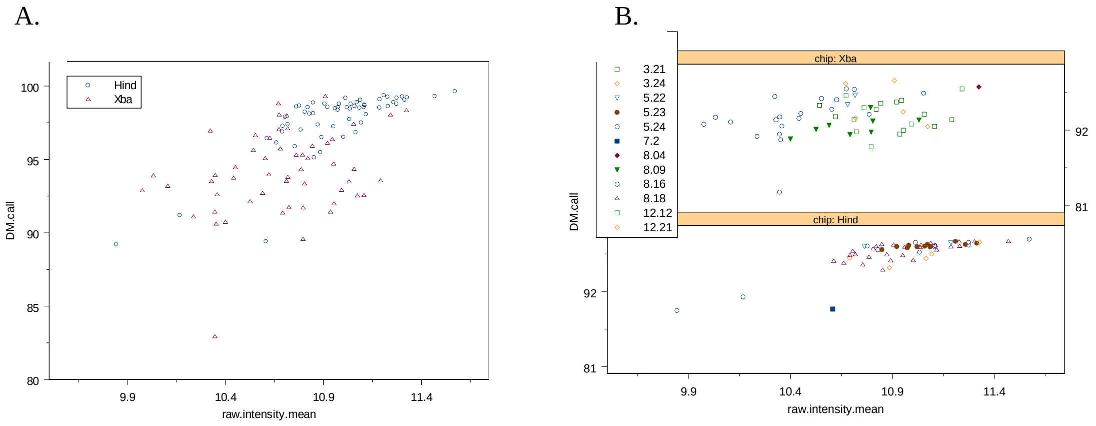


Figure 3. Overall chip intensity vs. DM call rate. The overall chip intensity for all samples was compared for both chip types. The greater variation in intensity among the XbaI digested samples shows the dependence of the call rate on chip intensity. A. Overall chip intensity of XbaI and HindIII versus the DM call rate. B. Hybridization date is shown by chip type.

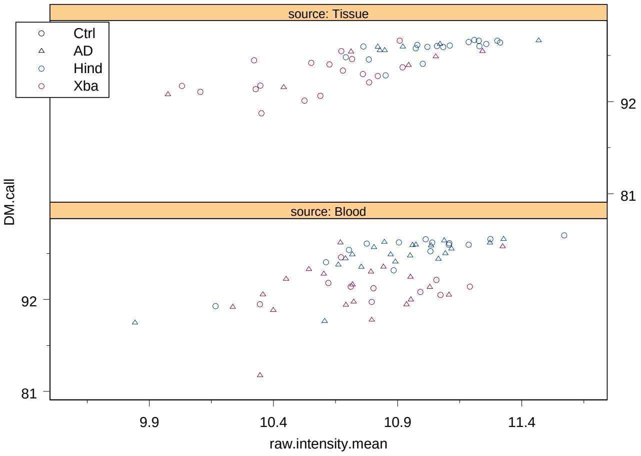


Figure 4. DNA source and phenotype had no effect on call rate.

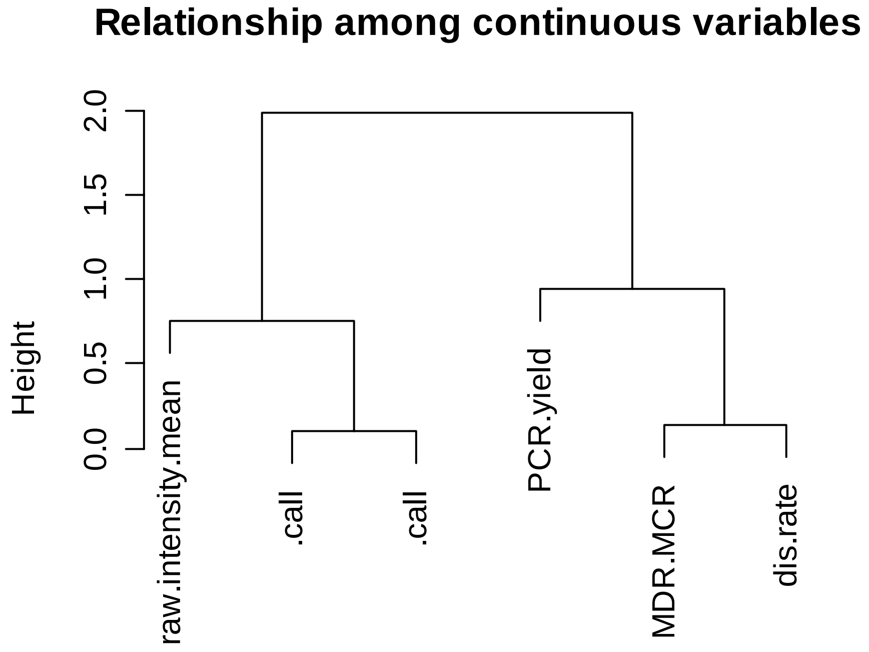


Figure 5. Clustering of sample characteristics. Continuous variables of all samples were clustered based on the distance metric (1-[correlation among all samples]).

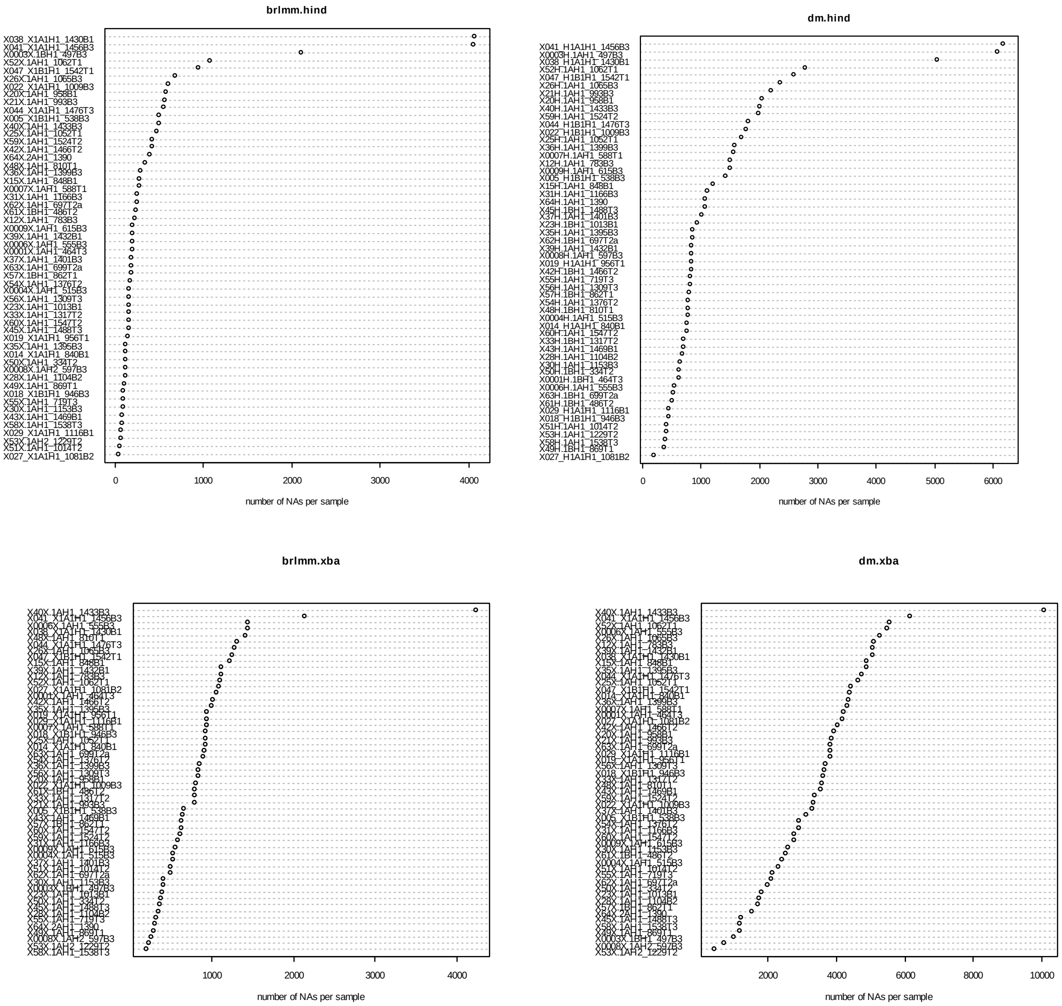
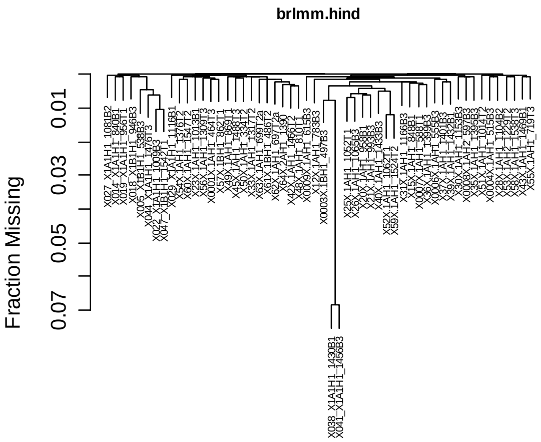
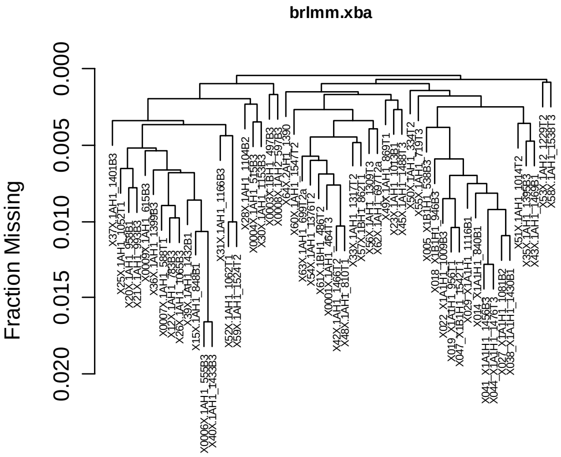


Figure6. Number of SNPs not called (NoCalls) per sample for each chip type and each genotyping algorithm.


brlmm.hind
Fraction Missing
0.01
0.03
0.05
0.07
X038_X1A1H1_1430B1
X041_X1A1H1_1456B3


brlmm.xba
Fraction Missing
0.000
0.005
0.010
0.015
0.020
X0006X1AH1_555B3
X40X1AH1_1433B3

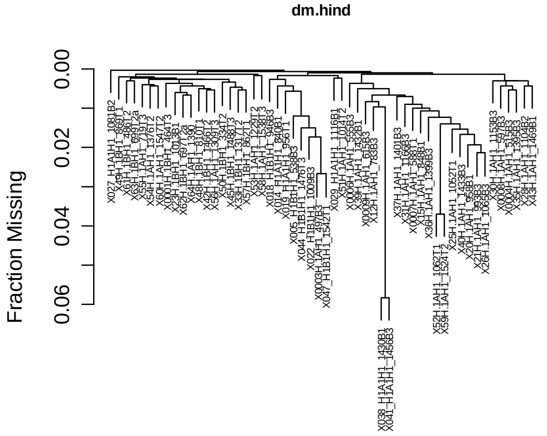
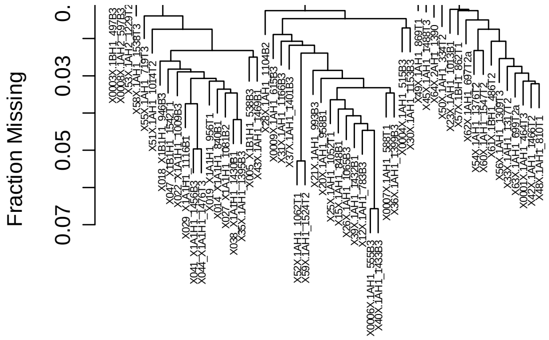
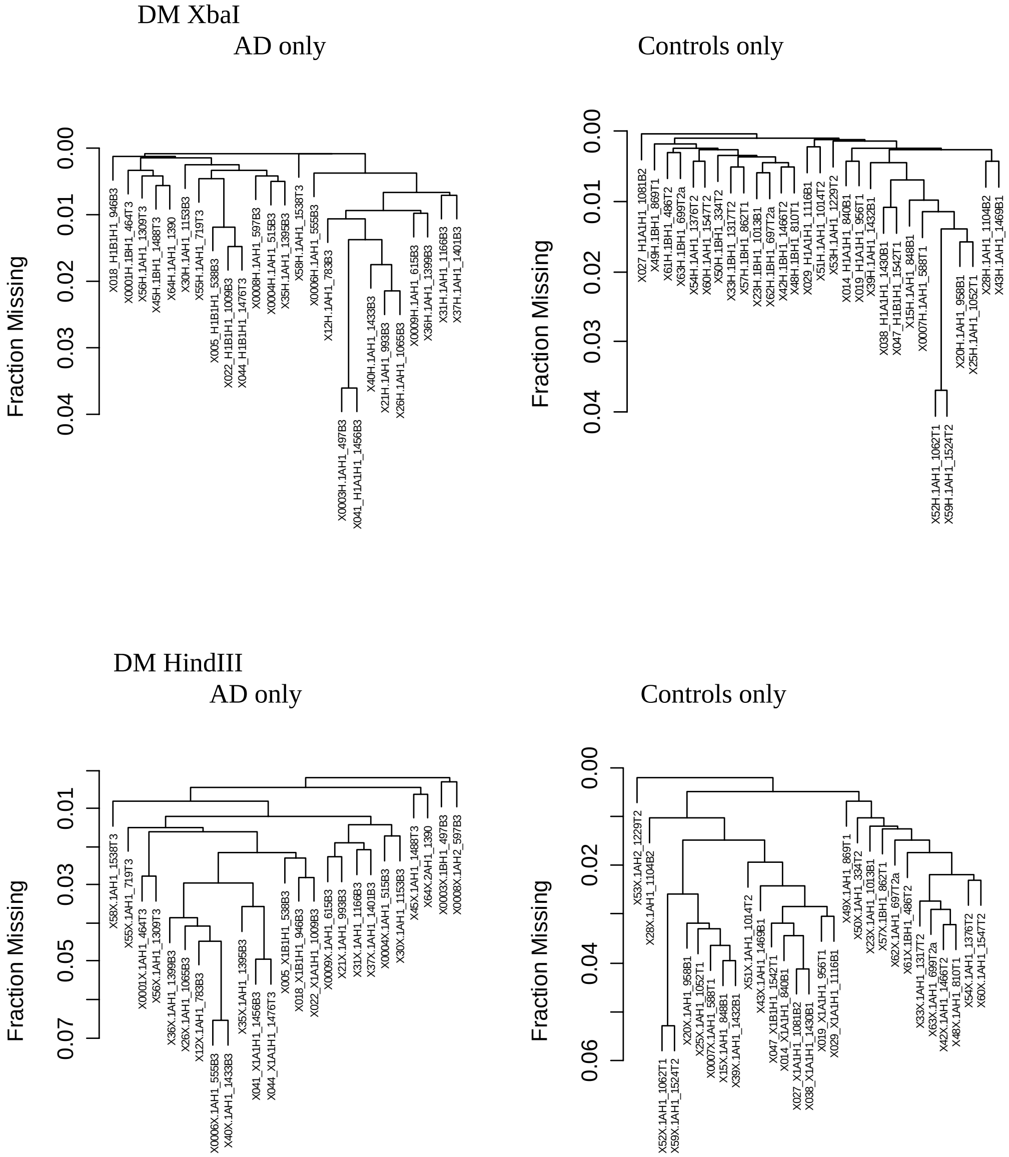


Fig. 7. Similarities among samples for SNPs which have NoCalls. The similarity measure is the fraction of NoCalls (missing data) in common between any two SNPs.





DM XbaI

AD only

Controls only

DM HindIII

AD only

Controls only

Figure 8. Similarities of NoCalls in SNPs within each phenotype.

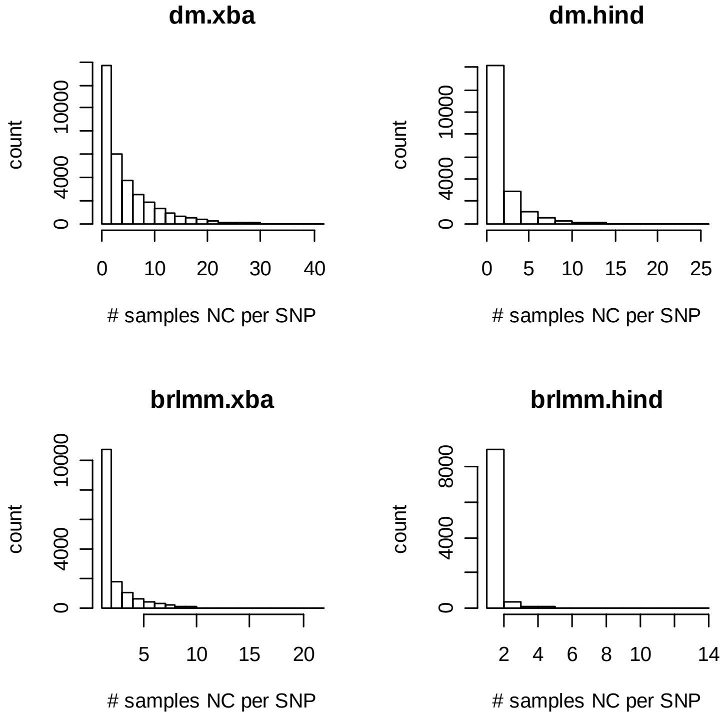


Figure 9. Distribution of the number of samples with a NoCall per SNP.

Chapter III. Tables.

Table 1. Parameters for Simulations

Parameter	Values
Intensity	
Chip	High, medium, low
Background	High, low
Standard Deviation	0.2, 0.5
Pixel Number	9, 20

Table 2. Summary of DM Model calls for AA and AB SNP simulations

A. All SNPs were simulated as AA

	intensity	Bck grd	SD	# pix	# SNPs simul	# called snps	# quartets called as highest maxn	# quartets called as highest maxaa	# quartets called as highest maxbb	# quartets called as highest maxab	concur
1	high	low	0.2	9	100	82	8	870	10	112	82
2	high	low	0.2	20	100	82	8	870	10	112	82
3	**high**	**low**	**0.5**	**9**	**100**	**84**	**8**	**866**	**9**	**117**	**84**
4	**high**	**low**	**0.5**	**20**	**100**	**84**	**8**	**866**	**9**	**117**	**84**
5	high	high	0.2	9	100	43	38	752	32	178	43
6	high	high	0.2	20	100	43	38	752	32	178	43
7	high	high	0.5	9	100	43	38	746	30	186	43
8	high	high	0.5	20	100	43	38	746	30	186	43
9	med	low	0.2	9	100	69	19	813	15	153	69
10	med	low	0.2	20	100	69	19	813	15	153	69
11	med	low	0.5	9	100	69	19	804	15	162	69
12	med	low	0.5	20	100	69	19	804	15	162	69
13	med	high	0.2	9	100	19	92	620	80	208	19
14	med	high	0.2	20	100	19	92	620	80	208	19
15	med	high	0.5	9	100	19	92	609	80	219	19
16	med	high	0.5	20	100	19	92	609	80	219	19
17	low	low	0.2	9	100	15	142	547	118	193	15
18	low	low	0.2	20	100	15	142	547	118	193	15
19	low	low	0.5	9	100	12	142	541	115	202	12
20	low	low	0.5	20	100	12	142	541	115	202	12
21	low	high	0.2	9	100	9	191	478	189	142	9
22	low	high	0.2	20	100	9	191	478	189	142	9
23	low	high	0.5	9	100	11	191	479	188	142	11
24	low	high	0.5	20	100	11	191	479	188	142	11

B. All SNPs were simulated as AB

	intensity	Bck grd	SD	# pix	# SNPs simul	# called snps	# quartets called as highest maxn	# quartets called as highest maxaa	# quartets called as highest maxbb	# quartets called as highest maxab	concur
1	**high**	**low**	**0.2**	**9**	**100**	**66**	**1**	**127**	**116**	**756**	**66**
2	**high**	**low**	**0.2**	**20**	**100**	**66**	**1**	**127**	**116**	**756**	**66**
3	high	low	0.5	9	100	72	1	110	107	782	72
4	high	low	0.5	20	100	72	1	110	107	782	72
5	high	high	0.2	9	100	24	14	180	210	596	24
6	high	high	0.2	20	100	24	14	180	210	596	24
7	high	high	0.5	9	100	31	14	167	192	627	31
8	high	high	0.5	20	100	31	14	167	192	627	31
9	med	low	0.2	9	100	38	11	174	152	663	38
10	med	low	0.2	20	100	38	11	174	152	663	38
11	med	low	0.5	9	100	46	11	162	136	691	46
12	med	low	0.5	20	100	46	11	162	136	691	46
13	med	high	0.2	9	100	5	65	277	240	418	5
14	med	high	0.2	20	100	5	65	277	240	418	5
15	med	high	0.5	9	100	8	65	265	224	446	8
16	med	high	0.5	20	100	8	65	265	224	446	8
17	low	low	0.2	9	100	0	105	284	280	331	0
18	low	low	0.2	20	100	0	105	284	280	331	0
19	low	low	0.5	9	100	1	105	286	275	334	1
20	low	low	0.5	20	100	1	105	286	275	334	1
21	low	high	0.2	9	100	0	141	334	315	210	0
22	low	high	0.2	20	100	0	141	334	315	210	0
23	low	high	0.5	9	100	0	141	331	314	214	0
24	low	high	0.5	20	100	0	141	331	314	214	0

Table 3. DM algorithm calls vs. summary score calls

A. AA Simulated genotypes

		Percent called correctly		Percent Summary Score		
Intensity	**background**	**DM algorithm**	**Summary score**	**Called BB**	**Called AB**	**NoCall**
hi	hi	43	98	0	0	0
hi	low	82	99	0	0	0
med	hi	19	85	0	7	1
med	low	69	100	0	0	0
low	hi	9	78	4	0	11
low	low	15	87	1	3	5

B. AB Simulated genotypes

		Percent called correctly		Percent Summary Score		
Intensity	**background**	**DM algorithm**	**Summary score**	**Called AA**	**Called BB**	**NoCall**
hi	hi	24	88	4	8	0
hi	low	66	100	0	0	0
med	hi	5	48	30	19	3
med	low	38	97	1	2	0
low	hi	0	12	48	30	10
low	low	0	35	32	28	5

Table 4. Variability among hybridizations of XbaI and HindIII SNP chips genotyped by the DM amd BRLMM algorithms.

A. XbaI chips	DM algorithm				BRLMM algorithm					
	call	**%AA**	**%AB**	**%BB**	**call**	**%AA**	**%AB**	**%BB**	**raw intensity mean[a]**	**raw intensity SD**
Mean	93.59	37.33	26.02	36.65	98.53	29.08	34.86	34.60	10.71	1.25
SD	3.22	1.01	1.69	0.76	1.01	0.54	0.59	0.85	0.31	0.09
CV	3.44	2.70	6.48	2.08	1.02	1.86	1.70	2.46	2.89	7.37
Lower95	92.82	37.08	25.62	36.47	98.27	28.94	34.71	34.38	10.63	1.23
Upper95	94.36	37.57	26.43	36.83	98.79	29.22	35.01	34.82	10.79	1.28
Min	82.92	35.64	19.78	35.48	92.81	28.01	32.05	29.59	9.98	1.09
Max	99.27	40.82	28.80	39.59	99.67	31.17	35.83	35.52	11.32	1.63
B. HindII chips										
Mean	97.73	38.03	24.79	37.18	99.26	26.03	36.97	36.26	10.96	1.34
SD	2.12	0.76	1.45	0.72	1.34	0.37	0.67	0.65	0.28	0.13
CV	2.17	1.99	5.87	1.93	1.35	1.41	1.82	1.79	2.58	9.69
Lower95	97.17	37.83	24.41	36.99	98.91	25.93	36.80	36.09	10.89	1.31
Upper95	98.28	38.22	25.17	37.37	99.61	26.12	37.15	36.43	11.04	1.38
Min	89.21	37.28	18.74	36.55	92.89	25.36	33.80	33.27	9.84	0.76
Max	99.64	41.32	26.17	40.27	99.92	27.08	37.58	36.92	11.57	1.59

[a] raw intensity of the entire chip

Table 5. Characteristics for all samples

Sample ID	chip	phenotype[a]	source	expr[b]	Chip Lot	Hybridization Date	gender
50X.1AH1.334T2	Xba	2	Tissue	N	4004432	5.24.2006	M
0001X.1AH1.464T3	Xba	3	Tissue	Y	4004432	5.24.2006	F
002.X1A1H1.486B2	Xba	2	Blood	Y	4004322	12.12.2005	F
61X.1BH1.486T2	Xba	2	Tissue	Y	4004432	5.22.2006	F
0004X.1AH1.515B3	Xba	3	Blood	N	4004432	3.21.2006	F
005.X1A1H1.538B3	Xba	3	Blood	N	4004322	12.12.2005	F
005.X1B1H1.538B3	Xba	3	Blood	N	4004432	12.21.2005	F
0008X.1AH1.597B3	Xba	3	Blood	N	4004432	3.21.2006	M
0008X.1AH2.597B3	Xba	3	Blood	N	4009296	3.24.2006	M
010.X1A1H1.697T2	Xba	1	Blood	Y	4004322	12.12.2005	M
010.X1B1H1.697T2	Xba	2	Blood	Y	4004432	12.21.2005	M
62X.1AH1.697T2a	Xba	2	Tissue	Y	4004432	5.24.2006	M
011.X1A1H1.699T1	Xba	1	Blood	Y	4004322	12.12.2005	M
63X.1AH1.699T2a	Xba	1	Tissue	Y	4004432	5.24.2006	M
55X.1AH1.719T3	Xba	3	Tissue	N	4004432	3.21.2006	M
48X.1AH1.810T1	Xba	1	Tissue	Y	4004432	5.24.2006	F
014.X1A1H1.840B1	Xba	3	Blood	N	4004322	12.12.2005	M
014.X1B1H1.840B1	Xba	1	Blood	N	4004432	12.21.2005	M
57X.1BH1.862T1	Xba	1	Tissue	N	4004432	5.22.2006	M
49X.1AH1.869T1	Xba	1	Tissue	Y	4004432	5.24.2006	M
018.X1A1H1.840B1	Xba	2	Blood	N	4004322	12.12.2005	M
018.X1B1H1.946B3	Xba	3	Blood	N	4004432	12.21.2005	M
019.X1A1H1.956T1	Xba	2	Blood	N	4004322	12.12.2005	M
022.X1A1H1.1009B3	Xba	1	Blood	N	4004322	12.12.2005	M
23X.1AH1.1013B1	Xba	1	Tissue	Y	4004432	5.24.2006	M
51X.1AH1.1014T2	Xba	2	Tissue	N	4004432	3.21.2006	M
027.X1A1H1.1081B2	Xba	1	Blood	N	4004322	12.12.2005	M
28X.1AH1.1104B2	Xba	2	Blood	N	4004432	3.21.2006	F
029.X1A1H1.1116B1	Xba	1	Blood	N	4004432	12.12.2005	M
30X.1AH1.1153B3	Xba	3	Blood	N	4004432	3.21.2006	M
53X.1AH1.1229T2	Xba	2	Tissue	Y	4004432	3.21.2006	F
53X.1AH2.1229T2	Xba	2	Tissue	Y	4009296	3.24.2006	F
56X.1AH1.1309T1	Xba	3	Tissue	Y	4004432	5.24.2006	F
33X.1AH1.1317T2	Xba	2	Tissue	Y	4004432	5.24.2006	F
54X.1AH1.1376T2	Xba	2	Tissue	N	4004432	3.21.2006	M
034.X1A1H1.1390T3	Xba	1	Blood	Y	4004432	12.12.2005	F
64X.1AH1.1390	Xba	3	Tissue	Y	4004432	5.24.2006	F
35X.1AH1.1395B3	Xba	3	Blood	N	4004432	3.21.2006	F
038.X1A1H1.1430B1	Xba	3	Blood	N	4004432	12.12.2005	F
041.X1A1H1.1456B3	Xba	3	Blood	N	4004432	12.12.2005	M
42X.1AH1.1466T2	Xba	2	Tissue	Y	4004432	5.24.2006	M
43X.1AH1.1469B1	Xba	1	Blood	N	4004432	3.21.2006	M
044.X1A1H1.1476T3	Xba	3	Blood	N	4004432	12.12.2005	M
45X.1AH1.1488T3	Xba	3	Tissue	Y	4004432	5.24.2006	M
58X.1AH1.1538T3	Xba	3	Tissue	N	4004432	3.21.2006	F
047.X1A1H1.1542T1	Xba	1	Blood	N	4004432	12.12.2005	F
047.X1B1H1.1542T1	Xba	1	Blood	N	4004432	12.21.2005	F
60X.1AH1.1547T2	Xba	2	Tissue	Y	4004432	3.21.2006	F

Sample ID	chip	phenotype[a]	source	expr[b]	Chip Lot	Hybridization Date	gender
50H.1BH1.334T2	Hind	2	Tissue	N	4004265	5.23.2006	M
0001H.1BH1.464T3	Hind	3	Tissue	Y	4004265	5.23.2006	F
002H.1A1H1.486B2	Hind	2	Blood	Y	4004265	2.2.2006	F
61H.1BH1.486T2	Hind	2	Tissue	Y	4004265	5.22.2006	F
0004H.1AH1.515B3	Hind	3	Blood	N	4004265	3.12.2006	F
005.H1B1H1.538B3	Hind	3	Blood	N	4004265	12.21.2005	F
0008H.1AH1.597B3	Hind	3	Blood	N	4004265	3.12.2006	M
010H.1A1H1.697T2	Hind	2	Blood	Y	4004265	2.2.2006	M
62H.1BH1.697T2a	Hind	2	Tissue	Y	4004265	5.23.2006	M
011H.1A1H1.699T1	Hind	1	Blood	Y	4004265	2.2.2006	M
63H.1BH1.699T2a	Hind	1	Tissue	Y	4004265	5.23.2006	M
55H.1AH1.719T3	Hind	3	Tissue	N	4004265	3.12.2006	M
48H.1BH1.810T1	Hind	1	Tissue	Y	4004265	5.23.2006	F
014H.1A1H1.840B1	Hind	1	Blood	N	4004265	2.2.2006	M
57H.1BH1.862T1	Hind	1	Tissue	N	4004265	5.22.2006	M
49H.1BH1.869T1	Hind	1	Tissue	Y	4004265	5.23.2006	M
018.H1B1H1.946B3	Hind	3	Blood	N	4004265	12.21.2005	M
019H.1A1H1.956T1	Hind	1	Blood	N	4004265	2.2.2006	M
022.H1B1H1.1009B3	Hind	3	Blood	N	4004265	12.21.2005	M
23H.1BH1.1013B1	Hind	1	Tissue	Y	4004265	5.23.2006	M
51H.1AH1.1014T2	Hind	2	Tissue	N	4004265	3.12.2006	M
027H.1A1H1.1081B2	Hind	2	Blood	N	4004265	2.2.2006	M
28H.1AH1.1104B2	Hind	2	Blood	N	4004265	3.12.2006	F
029H.1A1H1.1116B1	Hind	1	Blood	N	4004265	2.2.2006	M
30H.1AH1.1153B3	Hind	3	Blood	N	4004265	3.12.2006	M
53H.1AH1.1229T2	Hind	2	Tissue	Y	4004265	3.12.2006	F
56H.1BH1.1309T1	Hind	3	Tissue	Y	4004265	5.23.2006	F
33H.1BH1.1317T2	Hind	2	Tissue	Y	4004265	5.23.2006	F
54H.1AH1.1376T2	Hind	2	Tissue	N	4004265	3.12.2006	M
034H.1A1H1.1390T3	Hind	3	Blood	Y	4004265	2.2.2006	F
64H.1AH1.1390	Hind	3	Tissue	Y	4004265	5.24.2006	F
35H.1AH1.1395B3	Hind	3	Blood	N	4004265	3.12.2006	F
038H.1A1H1.1430B1	Hind	1	Blood	N	4004265	2.2.2006	F
041H.1A1H1.1456B3	Hind	3	Blood	N	4004265	2.2.2006	M
42H.1BH1.1466T2	Hind	2	Tissue	Y	4004265	5.23.2006	M
43H.1AH1.1469B1	Hind	1	Blood	N	4004265	3.12.2006	M
044.H1B1H1.1476T3	Hind	3	Blood	N	4004265	12.21.2005	M
45H.1BH1.1488T3	Hind	3	Tissue	Y	4004265	5.23.2006	M
58H.1AH1.1538T3	Hind	3	Tissue	N	4004265	3.12.2006	F
047.H1B1H1.1542T1	Hind	1	Blood	N	4004265	12.21.2005	F
60H.1AH1.1547T2	Hind	2	Tissue	Y	4004265	3.12.2006	F

Table 5. Characteristics for all samples, con't

Sample ID	AA DM	AB DM	BB DM	MCR[d]	MDR	MDR - MCR	PCR yield (ug)	brlmm call
50X.1AH1.334T2	37.02	26.34	36.64	96.8	99.83	3.03	69.39	99.47
0001X.1AH1.464T3	37.71	25.37	36.92	93.57	99.3	5.73	64.71	98.55
002.X1A1H1.486B2	36.73	26.97	36.3	93.19	98.91	5.72	56.1	98.96
61X.1BH1.486T2	36.63	27.26	36.11	96.35	99.67	3.32	47.46	98.88
0004X.1AH1.515B3	36.35	28.01	35.65	92.36	99.15	6.79	72.54	98.89
005.X1A1H1.538B3	37.96	24.87	37.16	86.42	97.1	10.68	68.9	97.35
005.X1B1H1.538B3	36.74	26.71	36.55	93	99.21	6.21	45.4	98.97
0008X.1AH1.597B3	37.78	25.61	36.61	85.16	96.73	11.57	74.07	97.24
0008X.1AH2.597B3	35.75	28.67	35.59	97.47	99.95	4.88	74.07	99.56
010.X1A1H1.697T2	37.61	24.77	37.63	84.88	97.15	12.27	65.4	96.86
010.X1B1H1.697T2	36.41	27.52	36.07	95.95	99.72	3.77	47.7	99.61
62X.1AH1.697T2a	36.75	26.94	36.31	96.4	99.71	3.31	45.09	99.32
011.X1A1H1.699T1	37.6	24.88	37.53	93	99.01	6.01	53.2	99.04
63X.1AH1.699T2a	37.32	25.86	36.82	92.95	98.91	5.96	58.95	98.7
55X.1AH1.719T3	36.21	27.72	36.06	95.62	99.54	3.92	62.19	99.4
48X.1AH1.810T1	36.69	27.17	36.14	95.02	99.11	4.09	56.7	97.93
014.X1A1H1.840B1	36.97	26.11	36.92	92.18	98.74	6.56	53.9	98.62
014.X1B1H1.840B1	39.28	21.68	39.03	83.78	96.27	12.49	44.6	95.81
57X.1BH1.862T1	36.74	26.9	36.35	97.35	99.76	2.41	49.05	99.15
49X.1AH1.869T1	35.89	28.06	36.05	97.89	99.83	1.94	65.52	99.57
018.X1A1H1.840B1	37.34	25.8	36.86	90.95	99.01	8.06	54.5	98.68
018.X1B1H1.946B3	37.03	26.13	36.84	92.02	99.17	7.15	41.9	98.69
019.X1A1H1.956T1	39.48	22.84	37.68	91.5	98.7	7.2	58.7	98.52
022.X1A1H1.1009B3	36.67	26.97	36.37	92.6	98.89	6.29	52.4	98.78
23X.1AH1.1013B1	36.48	27.34	36.18	96.77	99.8	3.03	62.37	99.43
51X.1AH1.1014T2	36.59	27.21	36.19	93.98	99.34	5.36	69.3	99.23
027.X1A1H1.1081B2	37.42	25.42	37.16	90.77	98.27	7.5	53.7	98.32
28X.1AH1.1104B2	36.21	27.94	35.85	94.62	99.66	5.04	70.74	99.28
029.X1A1H1.1116B1	37.25	25.97	36.78	92.63	99.06	6.43	54.2	98.59
30X.1AH1.1153B3	36.55	27.31	36.14	92.81	99.39	6.58	70.11	99.26
53X.1AH1.1229T2	37.04	26.7	36.26	93.2	99.26	6.06	63.09	98.75
53X.1AH2.1229T2	35.69	28.8	35.51	98.52	99.99	4.88	63.09	99.65
56X.1AH1.1309T1	36.94	26.12	36.94	93.56	99.19	5.63	60.66	98.8
33X.1AH1.1317T2	37.34	26.39	36.28	94.18	99.27	5.09	60.48	98.9
54X.1AH1.1376T2	36.94	26.83	36.23	94.07	99.16	5.09	59.94	98.7
034.X1A1H1.1390T3	37.94	25.02	37.03	92.04	98.79	6.75	52.7	98.55
64X.1AH1.1390	39.36	21.82	38.82	87	97.15	10.15	61.92	96.82
35X.1AH1.1395B3	38.13	24.88	36.99	87.64	97.68	10.04	65.16	98.12
038.X1A1H1.1430B1	38.35	24.29	37.37	87.14	97.62	10.48	55.7	97.68
041.X1A1H1.1456B3	37.36	25.93	36.71	89.04	97.58	8.54	48.5	96.76
42X.1AH1.1466T2	37.61	25.03	37.36	94.13	99.29	5.16	54	98.65
43X.1AH1.1469B1	37.6	25.92	36.48	91.4	98.85	7.45	66.33	98.92
044.X1A1H1.1476T3	37.74	25.3	36.96	90.05	98.5	8.45	49.3	98.04
45X.1AH1.1488T3	36.23	28.13	35.64	97.68	99.87	2.19	53.37	99.48
58X.1AH1.1538T3	36.04	28.48	35.48	96.74	99.81	3.07	63.54	99.64
047.X1A1H1.1542T1	37.71	25.07	37.22	89.57	98.46	8.89	55	98.1
047.X1B1H1.1542T1	37.34	25.22	37.44	91.88	98.79	6.91	53	98.24
60X.1AH1.1547T2	36.72	27.41	35.87	94.32	99.2	4.88	57.87	99.02

Sample ID	AA DM	AB DM	BB DM	MCR[d]	MDR	MDR - MCR	PCR yield (ug)	brlmm call
50H.1BH1.334T2	37.85	25.32	36.82	98.87	99.99	1.12	56.34	99.82
0001H.1BH1.464T3	37.5	25.62	36.89	98.62	99.99	1.37	52.16	99.8
002H.1A1H1.486B2	37.48	25.8	36.72	98.65	99.99	1.34	62.3	99.86
61H.1BH1.486T2	37.44	25.84	36.72	99.26	100	0.74	55.36	99.69
0004H.1AH1.515B3	37.7	25.49	36.81	97.89	99.93	2.04	70.7	99.68
005.H1B1H1.538B3	37.94	24.96	37.1	95.85	99.63	3.78	56.4	99.12
0008H.1AH1.597B3	37.59	25.49	36.92	98.19	99.91	1.72	81.6	99.77
010H.1A1H1.697T2	37.76	25.01	37.23	97.64	99.93	2.29	68.9	99.67
62H.1BH1.697T2a	37.8	24.77	37.43	98.38	99.99	1.61	54.99	99.7
011H.1A1H1.699T1	38.15	24.5	37.35	96.42	99.81	3.39	18.2	99.53
63H.1BH1.699T2a	37.7	25.35	36.95	98.97	99.99	1.02	54	99.78
55H.1AH1.719T3	37.69	25.26	37.05	98.62	99.94	1.32	59.8	99.84
48H.1BH1.810T1	37.74	25.45	36.8	98.65	99.97	1.32	55.8	99.58
014H.1A1H1.840B1	37.81	25.23	36.96	98.23	99.96	1.73	64.6	99.84
57H.1BH1.862T1	38.03	24.83	37.14	98.56	100	1.44	48.78	99.75
49H.1BH1.869T1	37.41	25.71	36.88	99.35	100	0.65	56.88	99.88
018.H1B1H1.946B3	37.28	26.17	36.55	98.72	99.96	1.24	46.5	99.85
019H.1A1H1.956T1	37.95	25.21	36.84	97.68	99.82	2.14	54.9	99.72
022.H1B1H1.1009B3	38.46	23.78	37.76	96.48	99.78	3.3	46.4	99.02
23H.1BH1.1013B1	38.03	24.9	37.07	98.48	99.96	1.48	52.47	99.79
51H.1AH1.1014T2	37.63	25.42	36.95	98.81	99.97	1.16	73.9	99.94
027H.1A1H1.1081B2	37.72	25.67	36.6	99.39	100	0.61	43.4	99.92
28H.1AH1.1104B2	37.48	25.88	36.64	98.31	99.96	1.65	62.3	99.79
029H.1A1H1.1116B1	37.32	25.91	36.77	98.8	99.97	1.17	48.8	99.9
30H.1AH1.1153B3	37.47	25.5	37.02	98.2	99.93	1.73	51.3	99.85
53H.1AH1.1229T2	37.78	25.5	36.72	99.15	99.96	0.81	61	99.9
56H.1BH1.1309T1	37.6	25.45	36.95	98.56	99.94	1.38	54.72	99.8
33H.1BH1.1317T2	37.54	25.7	36.76	98.68	100	1.32	54.27	99.8
54H.1AH1.1376T2	37.77	24.91	37.31	98.54	99.97	1.43	53.5	99.83
034H.1A1H1.1390T3	37.83	25.28	36.89	98.42	99.97	1.55	62.6	99.81
64H.1AH1.1390	37.96	25.03	37.01	97.92	99.93	2.01	57.15	99.5
35H.1AH1.1395B3	37.88	24.79	37.33	98.05	99.9	1.85	66.2	99.76
038H.1A1H1.1430B1	40.12	20.43	39.45	82.71	97.55	14.84	63.1	94.16
041H.1A1H1.1456B3	41.32	18.74	39.93	81.82	96.36	14.54	46.6	94.26
42H.1BH1.1466T2	38	24.79	37.21	98.5	99.97	1.47	50.13	99.47
43H.1AH1.1469B1	38.04	24.62	37.34	98.47	99.94	1.47	64.9	99.85
044.H1B1H1.1476T3	38.26	24.1	37.65	96.49	99.9	3.41	43	99.13
45H.1BH1.1488T3	38.2	24.34	37.46	98.14	99.94	1.8	57.78	99.81
58H.1AH1.1538T3	37.34	26.08	36.58	99.21	99.97	0.76	68.9	99.88
047.H1B1H1.1542T1	38.95	22.8	38.25	94.6	99.69	5.09	52	98.51
60H.1AH1.1547T2	37.58	25.66	36.76	98.71	99.99	1.28	56.7	99.82

Table 5. Characteristics for all samples, con't

Sample ID	AA brlmm	AB brlmm	BB brlmm	raw intensity mean	raw intensity SD
50X.1AH1.334T2	28.02	35.88	35.57	10.55	1.29
0001X.1AH1.464T3	29.15	34.88	34.51	9.98	1.2
002.X1A1H1.486B2	28.99	35.1	34.86	11.16	1.26
61X.1BH1.486T2	29.21	35.01	34.66	10.68	1.29
0004X.1AH1.515B3	29.45	34.87	34.57	10.84	1.26
005.X1A1H1.538B3	29.57	34	33.79	10.89	1.21
005.X1B1H1.538B3	29.06	34.86	35.04	10.95	1.36
0008X.1AH1.597B3	29.6	34.07	33.57	10.43	1.11
0008X.1AH2.597B3	28.9	35.38	35.27	10.67	1.39
010.X1A1H1.697T2	28.53	34.7	33.62	10.49	1.27
010.X1B1H1.697T2	28.64	35.62	35.35	11.27	1.14
62X.1AH1.697T2a	28.59	35.56	35.17	10.63	1.19
011.X1A1H1.699T1	28.95	35.08	35.01	11.08	1.27
63X.1AH1.699T2a	29.22	34.85	34.63	10.33	1.15
55X.1AH1.719T3	29.26	35.02	35.12	10.94	1.28
48X.1AH1.810T1	30.1	34.08	33.75	10.35	1.28
014.X1A1H1.840B1	28.71	34.83	35.08	11.11	1.23
014.X1B1H1.840B1	28.69	33.41	33.71	10.27	1.12
57X.1BH1.862T1	28.07	35.66	35.42	10.72	1.42
49X.1AH1.869T1	28.81	35.23	35.54	10.67	1.27
018.X1A1H1.840B1	28.95	34.71	35.01	10.96	1.3
018.X1B1H1.946B3	28.9	34.8	35	10.72	1.32
019.X1A1H1.956T1	28.38	35	35.14	11.19	1.2
022.X1A1H1.1009B3	28.68	34.96	35.15	11.06	1.24
23X.1AH1.1013B1	28.71	35.43	35.28	10.32	1.28
51X.1AH1.1014T2	28.69	35.3	35.24	10.92	1.37
027.X1A1H1.1081B2	29.35	34.5	34.47	10.99	1.23
28X.1AH1.1104B2	28.9	35.2	35.18	10.67	1.34
029.X1A1H1.1116B1	28.98	34.83	34.79	10.71	1.4
30X.1AH1.1153B3	29.2	35.01	35.05	10.54	1.29
53X.1AH1.1229T2	29.07	35.05	34.62	10.8	1.37
53X.1AH2.1229T2	28.83	35.48	35.34	10.91	1.63
56X.1AH1.1309T1	29.45	34.58	34.77	10.44	1.19
33X.1AH1.1317T2	29.59	35.05	34.26	10.03	1.14
54X.1AH1.1376T2	28.92	35.05	34.74	10.82	1.32
034.X1A1H1.1390T3	28.84	34.99	34.72	10.85	1.29
64X.1AH1.1390	28.91	34.06	33.86	9.6	1.08
35X.1AH1.1395B3	29.05	34.8	34.27	10.72	1.16
038.X1A1H1.1430B1	29.54	34.29	33.86	10.94	1.27
041.X1A1H1.1456B3	29.65	33.71	33.41	10.8	1.29
42X.1AH1.1466T2	28.64	35.05	34.95	10.11	1.22
43X.1AH1.1469B1	28.61	35.48	34.83	10.62	1.29
044.X1A1H1.1476T3	28.87	34.68	34.49	10.95	1.31
45X.1AH1.1488T3	28.87	35.5	35.11	10.71	1.33
58X.1AH1.1538T3	29.1	35.44	35.1	11.24	1.32
047.X1A1H1.1542T1	28.66	34.68	34.76	11.01	1.34
047.X1B1H1.1542T1	28.28	34.74	35.22	11.07	1.35
60X.1AH1.1547T2	29.56	34.97	34.5	10.76	1.27

Sample ID	AA brlmm	AB brlmm	BB brlmm	raw intensity mean	raw intensity SD
50H.1BH1.334T2	25.92	37.43	36.47	11.26	1.45
0001H.1BH1.464T3	26.29	37.05	36.45	11.07	1.44
002H.1A1H1.486B2	26.19	37.17	36.5	11.01	1.44
61H.1BH1.486T2	26.19	37.09	36.41	11.19	1.56
0004H.1AH1.515B3	26.26	37.12	36.3	11.04	1.28
005.H1B1H1.538B3	26.28	36.75	36.09	11.09	1.46
0008H.1AH1.597B3	26.35	37.03	36.39	10.97	1.23
010H.1A1H1.697T2	25.63	37.25	36.79	11.11	1.34
62H.1BH1.697T2a	25.61	37.22	36.87	11.08	1.36
011H.1A1H1.699T1	25.84	37.19	36.5	11.03	1.26
63H.1BH1.699T2a	25.85	37.34	36.6	11.31	1.49
55H.1AH1.719T3	26.06	37.18	36.6	10.82	1.37
48H.1BH1.810T1	26.2	37.12	36.26	11.06	1.46
014H.1A1H1.840B1	26.05	37.27	36.52	10.78	1.4
57H.1BH1.862T1	25.62	37.49	36.64	10.76	1.54
49H.1BH1.869T1	26.07	37.16	36.65	11.21	1.42
018.H1B1H1.946B3	26.55	37	36.3	11.33	1.48
019H.1A1H1.956T1	26.01	37.36	36.34	11.11	1.36
022.H1B1H1.1009B3	25.38	37.08	36.55	10.69	1.5
23H.1BH1.1013B1	25.82	37.42	36.56	10.97	1.38
51H.1AH1.1014T2	25.79	37.4	36.75	11.23	1.38
027H.1A1H1.1081B2	25.83	37.59	36.5	11.57	1.42
28H.1AH1.1104B2	26.56	37.02	36.21	10.91	1.3
029H.1A1H1.1116B1	26.36	37.04	36.51	11.27	1.4
30H.1AH1.1153B3	26.15	37.05	36.64	10.85	1.25
53H.1AH1.1229T2	25.85	37.52	36.53	11.3	1.39
56H.1BH1.1309T1	26.32	37.04	36.45	10.92	1.34
33H.1BH1.1317T2	26.47	37.05	36.28	10.98	1.48
54H.1AH1.1376T2	25.57	37.33	36.93	11.23	1.4
034H.1A1H1.1390T3	26.01	37.34	36.45	11.27	1.34
64H.1AH1.1390	26.02	37.19	36.3	10.83	1.35
35H.1AH1.1395B3	25.65	37.3	36.81	10.96	1.27
038H.1A1H1.1430B1	25.59	34.52	34.04	10.17	1.16
041H.1A1H1.1456B3	25.45	34.94	33.87	9.84	1.18
42H.1BH1.1466T2	25.64	37.27	36.56	11.02	1.53
43H.1AH1.1469B1	25.36	37.59	36.91	11.04	1.3
044.H1B1H1.1476T3	25.61	36.94	36.58	11.06	1.59
45H.1BH1.1488T3	25.47	37.51	36.83	10.85	1.4
58H.1AH1.1538T3	26.34	37.13	36.41	11.47	1.36
047.H1B1H1.1542T1	25.38	36.75	36.38	10.88	1.5
60H.1AH1.1547T2	26.37	37.11	36.34	11.11	1.35

[a] 1= Non-demented, low NFT, 2= Non-demented, mid tangles, 3= AD cases, high tangles

[b] Y = RNA from these samples were hybridized to HG-U133.Plus 2 expression arrays

[c] DM call = genotyping call rate from the Dynamic Modeling (DM) algorithm. AA DM is the homozygous AA only call rate.

[d]. %MCR is the call rate based on the MPAM algorithm. %MDR is the difference between the call rate and the detection rate as determined by the MPAM algorithm (Affymetrix).

Table 6. Number of discordant genotype calls between the DM and BRLMM algorithms

ID	sample	discordant rate	DM minimum call rate	# discordant calls	# discordant SNPs called by both programs	proportion discordant SNPs called by both programs	# calls AB/AA	# calls AB/BB
0001X.1AH1.464T3	464	7.51	92.49	4426	183	0.041	106	77
0003X.1BH1.497B3	497	1.94	98.06	1143	50	0.044	17	33
0004X.1AH1.515B3	515	4.25	95.75	2504	98	0.039	50	48
0006X.1AH1.555B3	555	9.91	90.09	5842	260	0.045	156	104
0007X.1AH1.588T1	588	7.63	92.37	4500	188	0.042	106	82
0008X.1AH2.597B3	597	1.35	98.65	798	21	0.026	10	11
0009X.1AH1.615B3	615	4.78	95.22	2817	100	0.035	53	47
005.X1B1H1.538B3	538	5.38	94.62	3174	91	0.029	47	44
014.X1A1H1.840B1	840	7.61	92.39	4485	157	0.035	90	67
018.X1B1H1.946B3	946	6.34	93.66	3736	125	0.033	72	53
019.X1A1H1.956T1	956	6.55	93.45	3860	131	0.034	76	54
022.X1A1H1.1009B3	1009	5.65	94.35	3332	115	0.035	60	55
027.X1A1H1.1081B2	1081	7.56	92.44	4457	201	0.045	121	79
029.X1A1H1.1116B1	1116	6.76	93.24	3984	161	0.040	90	71
038.X1A1H1.1430B1	1430	9.31	90.69	5487	302	0.055	175	127
041.X1A1H1.1456B3	1456	11.64	88.36	6860	355	0.052	200	155
044.X1A1H1.1476T3	1476	8.20	91.80	4833	204	0.042	124	80
047.X1B1H1.1542T1	1542	7.50	92.50	4422	109	0.025	57	52
12X.1AH1.783B3	783	9.17	90.83	5406	255	0.047	149	106
15X.1AH1.848B1	848	8.73	91.27	5146	221	0.043	114	107
20X.1AH1.958B1	958	6.82	93.18	4021	155	0.039	94	61
21X.1AH1.993B3	993	6.75	93.25	3977	148	0.037	74	74
23X.1AH1.1013B1	1013	3.15	96.85	1856	56	0.030	29	27
25X.1AH1.1052T1	1052	8.15	91.85	4806	191	0.040	122	68
26X.1AH1.1065B3	1065	9.49	90.51	5597	277	0.049	151	126
28X.1AH1.1104B2	1104	2.99	97.01	1764	60	0.034	35	25
30X.1AH1.1153B3	1153	4.38	95.62	2583	71	0.027	43	28
31X.1AH1.1166B3	1166	5.05	94.95	2978	90	0.030	49	41
33X.1AH1.1317T2	1317	6.32	93.68	3725	144	0.039	62	82
35X.1AH1.1395B3	1395	8.61	91.39	5079	197	0.039	113	84
36X.1AH1.1399B3	1399	7.62	92.38	4494	147	0.033	76	71
37X.1AH1.1401B3	1401	5.70	94.30	3359	128	0.038	73	55
39X.1AH1.1432B1	1432	8.91	91.09	5253	229	0.044	129	100
40X.1AH1.1433B3	1433	20.33	82.92	11986	805	0.067	443	362
42X.1AH1.1466T2	1466	7.11	92.89	4191	159	0.038	81	78
43X.1AH1.1469B1	1469	6.06	93.94	3573	106	0.030	57	49
45X.1AH1.1488T3	1488	2.13	97.87	1257	48	0.038	28	20
48X.1AH1.810T1	810	6.77	93.23	3989	194	0.049	98	96
49X.1AH1.869T1	869	2.04	97.96	1200	37	0.031	20	17
50X.1AH1.334T2	334	3.36	96.64	1982	62	0.031	32	30
51X.1AH1.1014T2	1014	3.81	96.19	2245	63	0.028	30	33
52X.1AH1.1062T1	1062	9.68	90.32	5710	240	0.042	142	98
53X.1AH2.1229T2	1229	0.89	99.11	526	29	0.055	17	12
54X.1AH1.1376T2	1376	5.35	94.65	3157	129	0.041	76	53
55X.1AH1.719T3	719	3.62	96.38	2137	62	0.029	32	30
56X.1AH1.1309T1	1309	6.64	93.36	3912	142	0.036	83	59

57X.1BH1.862T1	862	3.10	96.90	1826	59	0.032	35	24
58X.1AH1.1538T3	1538	1.94	98.06	1144	28	0.024	14	14
59X.1AH1.1524T2	1524	5.79	94.21	3415	159	0.047	93	66
60X.1AH1.1547T2	1547	4.88	95.12	2877	113	0.039	74	39
61X.1BH1.486T2	486	4.58	95.42	2698	89	0.033	43	46
62X.1AH1.697T2a	697	3.61	96.39	2127	53	0.025	25	28
63X.1AH1.699T2a	699	6.85	93.15	4039	135	0.033	76	59
64X.2AH1.1390	1390	2.72	97.28	1601	44	0.027	23	21

Table 7. Chisquare Test of the number of NoCalls across all samples per chromosome. (p-value = 0.001)

	observed		**expected**	
chr	**≥ 10**	**≤ 5**	**≥ 10**	**≤ 5**
1	459	1611	467	1603
2	546	1835	538	1843
3	355	1327	380	1302
4	420	1432	418	1434
5	466	1452	433	1485
6	497	1420	433	1484
7	389	1267	374	1282
8	345	1166	341	1170
9	215	829	236	808
10	296	1019	297	1018
11	276	1005	289	992
12	265	970	279	956
13	260	907	263	904
14	232	694	209	717
15	149	514	150	513
16	85	421	114	392
17	71	341	93	319
18	193	631	186	638
19	37	126	37	126
20	102	359	104	357
21	110	353	105	358
22	23	112	30	105
23	102	417	117	402

Table 8. Chisquare test of the number of NoCall SNPs relative to PCR fragment length

A. DM algorithm

p-value = 2.2 x 10^{-16}

	Number of SNPs within each bin					
	Observed [b]			**Expected**		
	# NC across samples			**# NC across samples**		
Fragment Length Bins[a]	**0**	**≥ 10**	**≤ 5**	**0**	**≥ 10**	**≤ 5**
(198,439]	871	115	537	837	155	532
(439,680]	6187	688	3421	5656	1046	3594
(680,920]	6137	536	3254	5453	1008	3466
(920,1160]	6183	705	3476	5693	1053	3618
(1160,1400]	5436	888	3596	5449	1008	3463
(1400,1640]	3946	1072	3100	4459	825	2834
(1640,1880]	2491	1268	2260	3306	611	2101
(1880,2120]	873	668	772	1271	235	807

B. BRLMM algorithm

p-value = 2.2 x 10^{-16}

	Number of SNPs within each bin					
	Observed			**Expected**		
	# NC across samples			**# NC across samples**		
Fragment Length Bins	**0**	**≥ 10**	**≤ 5**	**0**	**≥ 10**	**≤ 5**
(198,439]	1167	14	342	1114	10	399
(439,680]	8167	60	2069	7529	68	2699
(680,920]	7980	45	1902	7259	66	2602
(920,1160]	8132	44	2188	7578	69	2717
(1160,1400]	7411	44	2465	7254	66	2600
(1400,1640]	5472	57	2589	5936	54	2128
(1640,1880]	3359	78	2582	4401	40	1578
(1880,2120]	1073	47	1193	1691	15	606

[a] PCR fragment lengths were divided into bins of 240 bp

[b] Number of NoCall samples in each bin with 0, ≥ 10 NoCalls, or ≤5 NoCalls

Table 9. Chisquare test of the numbers of calls per genotype for all SNPs for 4 individual samples

Sample 1433B3; DM call rate = 82.92; BRLMM call rate = 92.81; overall chip intensity = 10.34

A. DM algorithm

p-value = 2.2 x 10^{-16}

	Observed # of each genotype				Expected # of each genotype			
Bins[a]	AA	AB	BB	NC	AA	AB	BB	NC
(198,439]	475	330	426	292	515	279	469	260
(439,680]	3402	2082	3194	1618	3485	1883	3170	1758
(680,920]	3333	1960	3169	1465	3360	1816	3056	1695
(920,1160]	3528	1970	3184	1682	3508	1896	3191	1769
(1160,1400]	3422	1739	3107	1652	3358	1815	3054	1694
(1400,1640]	2701	1399	2467	1551	2748	1485	2499	1386
(1640,1880]	2120	895	1751	1253	2037	1101	1853	1028
(1880,2120]	813	322	707	471	783	423	712	395

B. BRLMM algorithm

p-value = 3.58 x 10^{-6}

	Observed # of each genotype				Expected # of each genotype			
Bins	AA	AB	BB	NC	AA	AB	BB	NC
(198,439]	115	446	547	415	109	488	474	451
(439,680]	707	3217	3355	3017	739	3300	3204	3052
(680,920]	668	3143	3083	3033	713	3182	3089	2943
(920,1160]	764	3314	3210	3076	744	3322	3225	3072
(1160,1400]	791	3190	3001	2938	712	3180	3087	2941
(1400,1640]	554	2574	2592	2398	583	2602	2526	2407
(1640,1880]	416	2070	1787	1746	432	1929	1873	1784
(1880,2120]	185	792	623	713	166	741	720	686

Sample 1062T1; DM call rate = 90.59; BRLMM call rate = 98.13; overall chip intensity = 10.35

A. DM algorithm

p-value = 2.2 x 10^{-16}

	Observed # of each genotype				Expected # of each genotype			
Bins	AA	AB	BB	NC	AA	AB	BB	NC
(198,439]	493	384	458	188	531	341	508	143
(439,680]	3559	2425	3326	986	3592	2303	3432	970
(680,920]	3399	2387	3346	795	3463	2220	3309	935
(920,1160]	3610	2400	3530	824	3615	2318	3454	976
(1160,1400]	3521	2151	3329	919	3460	2219	3306	934
(1400,1640]	2806	1715	2754	843	2832	1816	2706	765
(1640,1880]	2177	1185	1973	684	2100	1346	2006	567
(1880,2120]	835	433	776	269	807	517	771	218

B. BRLMM algorithm

p-value = 2.4 x 10^{-9}

	Observed # of each genotype				Expected # of each genotype			
Bins	AA	AB	BB	NC	AA	AB	BB	NC
(198,439]	51	498	508	466	29	532	443	519
(439,680]	239	3554	3134	3369	193	3599	2996	3508
(680,920]	156	3402	2956	3413	186	3470	2889	3383
(920,1160]	191	3604	2995	3574	194	3622	3016	3532
(1160,1400]	180	3528	2812	3400	186	3467	2887	3380
(1400,1640]	136	2813	2327	2842	152	2837	2362	2766
(1640,1880]	101	2202	1665	2051	113	2104	1752	2051
(1880,2120]	41	839	621	812	43	808	673	788

[a] PCR fragment lengths were divided into bins of 240 bp

Table 9. Chisquare test of the numbers of calls per genotype for all SNPs for 4 individual samples, con’t

Sample 1081B2; DM call rate = 92.89; BRLMM call rate = 98.16; overall chip intensity = 10.99

A. DM algorithm

p-value = 2.2 x 10^{-16}

	Observed # of each genotype				Expected # of each genotype			
Bins[a]	AA	AB	BB	NC	AA	AB	BB	NC
(198,439]	500	426	499	98	528	365	522	108
(439,680]	3538	2776	3408	574	3573	2465	3526	732
(680,920]	3459	2607	3422	439	3445	2377	3400	706
(920,1160]	3625	2623	3553	563	3596	2481	3550	737
(1160,1400]	3479	2273	3516	652	3442	2375	3398	705
(1400,1640]	2803	1752	2847	716	2817	1943	2780	577
(1640,1880]	2106	1140	2024	749	2089	1441	2062	428
(1880,2120]	783	403	761	366	803	554	792	164

B. BRLMM algorithm

p-value = 2.2 x 10^{-16}

	Observed # of each genotype				Expected # of each genotype			
Bins	AA	AB	BB	NC	AA	AB	BB	NC
(198,439]	30	493	502	498	28	524	447	524
(439,680]	145	3517	3219	3415	189	3540	3024	3543
(680,920]	109	3443	2944	3431	183	3413	2915	3416
(920,1160]	142	3603	3052	3567	191	3563	3044	3566
(1160,1400]	170	3443	2775	3532	183	3411	2913	3413
(1400,1640]	202	2754	2317	2845	149	2791	2384	2793
(1640,1880]	189	2081	1701	2048	111	2069	1768	2071
(1880,2120]	89	773	665	786	43	795	679	796

Sample 1390T3; DM call rate = 97.39; BRLMM call rate = 99.5; overall chip intensity = 11.05

A. DM algorithm

p-value = 2.2 x 10^{-16}

	Observed # of each genotype				Expected # of each genotype			
Bins	AA	AB	BB	NC	AA	AB	BB	NC
(198,439]	490	517	507	9	535	413	535	40
(439,680]	3537	3108	3541	110	3619	2795	3613	269
(680,920]	3455	2872	3512	88	3489	2695	3484	259
(920,1160]	3682	2907	3648	127	3642	2814	3637	270
(1160,1400]	3510	2663	3569	178	3486	2693	3481	259
(1400,1640]	2858	2047	2912	301	2853	2204	2849	212
(1640,1880]	2179	1347	2043	450	2115	1634	2112	157
(1880,2120]	842	416	792	263	813	628	812	60

B. BRLMM algorithm

p-value = 2.2 x 10^{-16}

	Observed # of each genotype				Expected # of each genotype			
Bins	AA	AB	BB	NC	AA	AB	BB	NC
(198,439]	6	491	520	506	8	537	441	538
(439,680]	16	3549	3175	3556	51	3628	2980	3636
(680,920]	12	3461	2925	3529	49	3498	2874	3506
(920,1160]	28	3690	2986	3660	51	3652	3000	3660
(1160,1400]	24	3521	2799	3576	49	3496	2872	3504
(1400,1640]	52	2862	2263	2941	40	2861	2350	2867
(1640,1880]	90	2187	1662	2080	30	2121	1742	2126
(1880,2120]	61	848	598	806	11	815	670	817

[a] PCR fragment lengths were divided into bins of 240 bp

Table 10. Proportion of calls of all SNPs across all samples

		Average proportion of all SNPs			Total number of SNPs	
		hetAA	hetAB	hetBB	all samples AB hets	invariant across all samples
DM	**XbaI**	0.3704	0.2639	0.3657	9	4399
DM	**HindIII**	0.3797	0.2489	0.3714	13	9483
BRLMM	**XbaI**	0.3535	0.2952	0.3513	9	5421
BRLMM	**HindIII**	0.3724	0.2624	0.3652	13	8869

Table 11. SNPs significantly out of Hardy-Weinberg Equilibrium

A. SNPs for which all samples were used in calculating HWE

		unadjusted p-value for samples:			FDR	
		all	AD	ctrl	ctrl	AD
DM.Hind	Minimum p-val	8.03E-16	2.65E-07	1.79E-08	0.000102213	0.0012662
	number of SNPs with p-val <0.05	2970	1883	1942	107	47
DM.Xba	Minimum p-val	8.03E-16	4.28E-08	1.79E-08	0.000216225	0.0022358
	number of SNPs with p-val <0.05	5593	3160	3639	232	109
BRLMM.Hind	Minimum p-val	8.03E-16	2.65E-07	1.79E-08	0.0000852	0.0009497
	number of SNPs with p-val <0.05	2268	1485	1661	83	44
BRLMM.Xba	Minimum p-val	8.03E-16	4.28E-08	1.79E-08	0.000131596	0.0014228
	number of SNPs with p-val <0.05	3024	1878	2160	89	60

B. SNPs with 0, 1 or 2 NoCalls across all samples

		unadjusted p-value for samples:			FDR	
		all	AD	ctrl	ctrl	AD
DM.Hind	Minimum p-val	8.03E-16	2.65E-07	1.79E-08	0.0002162	0.001266
	number of SNPs with p-val <0.05	1919	1240	1362	57	25
DM.Xba	Minimum p-val	8.03E-16	4.28E-08	1.79E-08	0.0002162	0.002236
	number of SNPs with p-val <0.05	1767	1126	1287	70	34
BRLMM.Hind	Minimum p-val	8.03E-16	2.65E-07	1.79E-08	0.0000851	0.000978
	number of SNPs with p-val <0.05	2130	1423	1580	69	35
BRLMM.Xba	Minimum p-val	8.03E-16	4.28E-08	1.79E-08	0.0001423	0.0012758
	number of SNPs with p-val <0.05	2431	1541	1810	65	40

Table 12. Relationship between SNPs out of HWE and fragment length

A. HWE all uncorrected p-val

p-value = 2.2 x 10 $^{-16}$

	observed		expected	
Frag. length[a]	**p>0.05**	**p<0.05**	**p>0.05**	**p<0.05**
(198,439]	1393	130	1380	143
(439,680]	9414	882	9331	965
(680,920]	9189	738	8997	930
(920,1160]	9499	865	9393	971
(1160,1400]	9026	894	8991	929
(1400,1640]	7259	859	7357	761
(1640,1880]	5208	811	5455	564
(1880,2120]	2013	300	2096	217

B. HWE corrected p-val for controls

p-value = 0.00074

	observed		expected	
Frag. length	**p>0.05**	**p<0.05**	**p>0.05**	**p<0.05**
(198,439]	1520	3	1518	5
(439,680]	10265	31	10260	36
(680,920]	9909	18	9893	34
(920,1160]	10337	27	10328	36
(1160,1400]	9874	46	9886	34
(1400,1640]	8087	31	8090	28
(1640,1880]	5987	32	5998	21
(1880,2120]	2299	14	2305	8

C. HWE corrected p-value for AD cases

p-value = 0.9432

	observed		expected	
Frag. length	**p>0.05**	**p<0.05**	**p>0.05**	**p<0.05**
(198,439]	1521	2	1521	2
(439,680]	10281	15	10281	15
(680,920]	9915	12	9912	15
(920,1160]	10350	14	10349	15
(1160,1400]	9903	17	9905	15
(1400,1640]	8105	13	8106	12
(1640,1880]	6010	9	6010	9
(1880,2120]	2309	4	2310	3

[a] PCR fragment lengths were divided into bins of 240 bp and the number of SNPs out of HWE (p<0.05) was compared among bins by χ^2 analysis.

Table 13. Relationship of PCR fragment length and Average Intensity values per SNP.

Sample 1081B2; DM call rate = 92.89; BRLMM call rate = 98.16; overall chip intensity = 10.99

p-value = 2.2 x 10^{-16}

	Observed					**Expected**				
Fragment	**Intensity value bins[b]**					**Intensity value bins**				
Length Bins[a]	**[814, 1941)**	**[1941, 2417)**	**[2417, 2975)**	**[2975, 3852)**	**[3852,16301]**	**[814, 1941)**	**[1941, 2417)**	**[2417, 2975)**	**[2975, 3852)**	**[3852,16301]**
(198,439]	210	283	317	354	359	305	305	305	305	304
(439,680]	1189	1529	1988	2408	3182	2061	2063	2060	2060	2053
(680,920]	976	1369	1871	2477	3234	1987	1989	1986	1986	1979
(920,1160]	1355	1808	2206	2450	2545	2074	2077	2074	2073	2066
(1160,1400]	1955	2171	2213	2060	1521	1986	1988	1985	1984	1978
(1400,1640]	2390	2137	1716	1257	618	1625	1627	1624	1624	1618
(1640,1880]	2491	1741	1034	576	177	1205	1206	1204	1204	1200
(1880,2120]	1139	679	357	116	22	463	463	463	463	461

Sample 1062T1; DM call rate = 90.59; BRLMM call rate = 98.13; overall chip intensity = 10.35

p-value = 2.2 x 10^{-16}

	Observed					**Expected**				
Fragment	**Intensity value bins**					**Intensity value bins**				
Length Bins	**[814, 1941)**	**[1941, 2417)**	**[2417, 2975)**	**[2975, 3852)**	**[3852,16301]**	**[814, 1941)**	**[1941, 2417)**	**[2417, 2975)**	**[2975, 3852)**	**[3852,16301]**
(198,439]	244	273	298	333	375	305	305	305	305	303
(439,680]	1586	1706	1898	2202	2904	2063	2061	2060	2061	2051
(680,920]	1443	1621	1876	2200	2787	1989	1987	1986	1987	1978
(920,1160]	1768	1921	2123	2235	2317	2077	2075	2073	2074	2065
(1160,1400]	2091	2074	2062	2049	1644	1988	1986	1984	1986	1976
(1400,1640]	2073	1930	1697	1431	987	1627	1625	1624	1625	1617
(1640,1880]	1759	1579	1269	934	478	1206	1205	1204	1205	1199
(1880,2120]	756	603	475	321	158	464	463	463	463	461

[a] PCR fragment lengths were divided into bins of 240 bp

[b] Intensity values of SNPs were divided into bins and the number of SNPs in each bin was compared by χ^2 analysis

Table 13. Relationship of PCR fragment length and Average Intensity values per SNP, con't.

Sample 1433B3; DM call rate = 82.92; BRLMM call rate = 92.81; overall chip intensity = 10.34

p-value = 2.2 x 10^{-16}

	Observed					**Expected**				
Fragment	**Intensity value bins[b]**					**Intensity value bins**				
Length Bins[a]	**[814, 1941)**	**[1941, 2417)**	**[2417, 2975)**	**[2975, 3852)**	**[3852,16301]**	**[814, 1941)**	**[1941, 2417)**	**[2417, 2975)**	**[2975, 3852)**	**[3852,16301]**
(198,439]	241	240	333	340	369	305	305	305	305	303
(439,680]	1605	1725	1926	2265	2775	2064	2060	2061	2059	2052
(680,920]	1494	1751	1998	2193	2491	1990	1986	1987	1986	1978
(920,1160]	1838	2010	2091	2241	2184	2078	2074	2074	2073	2065
(1160,1400]	2097	2069	2077	1933	1744	1989	1985	1985	1984	1977
(1400,1640]	2035	1852	1644	1413	1174	1627	1624	1625	1624	1618
(1640,1880]	1705	1495	1187	969	663	1207	1204	1205	1204	1199
(1880,2120]	709	560	448	343	253	464	463	463	463	461

Sample 1390T3; DM call rate = 97.39; BRLMM call rate = 99.5; overall chip intensity = 11.05

p-value = 2.2 x 10^{-16}

	Observed					**Expected**				
Fragment	**Intensity value bins**					**Intensity value bins**				
Length Bins	**[814, 1941)**	**[1941, 2417)**	**[2417, 2975)**	**[2975, 3852)**	**[3852,16301]**	**[814, 1941)**	**[1941, 2417)**	**[2417, 2975)**	**[2975, 3852)**	**[3852,16301]**
(198,439]	49	125	233	373	743	305	305	305	305	303
(439,680]	506	1072	1792	2706	4220	2062	2063	2060	2059	2051
(680,920]	531	1130	1997	2727	3542	1988	1989	1986	1986	1978
(920,1160]	897	1910	2530	2855	2172	2076	2077	2074	2073	2065
(1160,1400]	1798	2571	2734	2044	773	1987	1988	1985	1984	1976
(1400,1640]	2794	2651	1718	786	169	1626	1627	1624	1624	1617
(1640,1880]	3460	1785	584	161	29	1206	1206	1204	1204	1199
(1880,2120]	1678	474	113	45	3	463	463	463	463	461

[a] PCR fragment lengths were divided into bins of 240 bp

[b] Intensity values of SNPs were divided into bins and the number of SNPs in each bin was compared by χ^2 analysis

Table 14. Relationship between SNPs out of HWE and Number of NoCalls across samples

A. HWE all uncorrected p-val

p-value = 2.2 x 10 $^{-16}$

	Observed		Expected	
# NC across samples[a]	**p>0.05**	**p<0.05**	**p>0.05**	**p<0.05**
(0,5]	21270	1626	19855	3041
(5,10]	5522	1280	5899	903
(10,15]	2075	879	2562	392
(15,20]	728	494	1060	162
(20,25]	324	233	483	74
(25,30]	132	70	175	27
(30,35]	69	30	86	13
(35,40]	21	3	21	3
(40,47]	9	2	10	1

B. HWE corrected p-val for controls

p-value = 2.2 x 10 $^{-16}$

	Observed		Expected	
# NC across samples	**p>0.05**	**p<0.05**	**p>0.05**	**p<0.05**
(0,5]	22827	69	22766	130
(5,10]	6748	54	6763	39
(10,15]	2918	36	2937	17
(15,20]	1202	20	1215	7
(20,25]	541	16	554	3
(25,30]	200	2	201	1
(30,35]	99	0	98	1
(35,40]	24	0	24	0
(40,47]	11	0	11	0

C. HWE corrected p-value for AD cases

p-value = 2.9 x 10 $^{-12}$

	Observed		Expected	
# NC across samples	**p>0.05**	**p<0.05**	**p>0.05**	**p<0.05**
(0,5]	22868	28	22835	61
(5,10]	6775	27	6784	18
(10,15]	2931	23	2946	8
(15,20]	1212	10	1219	3
(20,25]	553	4	556	1
(25,30]	202	0	201	1
(30,35]	99	0	99	0
(35,40]	24	0	24	0
(40,47]	11	0	11	0

[a] The number of NoCalls was divided into bins and the number of SNPs out of HWE (p<0.05) was compared among bins by χ^2 analysis.

Chapter IV. The Effect of Copy Number Variation on Gene Expression in Age-related Cognitive Decline

Introduction:

Techniques for studying genome wide phenomena have revealed the complexity of cellular processes and the underlying genetic variation contributing to them. The appreciation of the dynamic nature of the human genome has expanded our view of the ways in which DNA sequence variation can contribute to phenotypic differences. Genomic structural variation from 1 base pair to many megabases in size have been associated with various disease processes [43]. Gene expression levels in particular have been shown to be affected by the more subtle variations in DNA sequence such as SNPs [3, 38, 122, 123]. Larger variations in DNA structure such as amplifications and deletions of chromosomal DNA have been known to affect gene expression levels in cancer cells [52-55] and have recently been shown to affect expression in normal individuals[56]. However, it is unknown how individual copy number variants may impact differential gene expression as determined in whole genome profiling studies from studies investigating diseases other than cancer.

We set out to investigate the influence of normal copy number variation (CNV) on our gene expression study (Chapter II, [57]) using Affymetrix 100K GeneChips. Because the majority of copy number studies concern cancer cells, the algorithms for estimating copy number are optimized for the particular genomic characteristics of cancer DNA anomalies. We evaluated several algorithms with respect to the hybridization assay protocol and characteristics of the DNA fragments hybridized to the chips (Chapter III) to determine the most robust algorithm for detecting copy number variation across samples of non-cancer cells. We highlight the algorithmic factors necessary for effective normalization across arrays in an experiment and the effect it has on the ability of the various algorithms to detect and estimate CNVs. We found that both common copy number variation as well as the individual copy number variation among the samples could contribute to differential gene expression in gene profiling studies.

2. Materials and Methods:

2A. DNA isolation and Hybridization:

DNA from subjects with a clinical diagnosis of no dementia within a year of death (N= 30) or Probable AD (N= 25) [65]was used in this study. All subjects were participants in studies performed by the Layton Aging and Alzheimer 's disease Center, Portland, Oregon which maintains an extensive collection of well characterized sample material and clinical data. DNA was isolated from either whole blood or postmortem human frontal cortex brain tissue. DNA from whole blood was isolated using the QIAmp DNA blood kit (Qiagen, Valencia, CA). For deceased subjects, approximately 100 mg of brain tissue (previously frozen at -80^{o}C) was processed for genomic DNA using the Wizard Genomic DNA purification Kit (Promega, Madison WI) following manufacturer's instructions.

Isolated genomic DNA from each subject was digested and labeled following manufacturer's instructions (Affymetrix Inc., Santa Clara, CA). Briefly, 250ng of genomic DNA was digested with a restriction enzyme (XbaI or HindIII), ligated to an appropriate adapter for each enzyme, and amplified by PCR using a single primer. The PCR products were then digested with DNaseI, labeled and hybridized separately to the Affymetrix GeneChip Mapping 100K array chips. The arrays were scanned and genotypes called by the DM and BRLMM algorithms. These arrays contain probe sets to interrogate 58960 (XbaI) and 58974 (HindIII) SNPs across the entire human genome. Sample labeling and array hybridizations and processing were performed in the Affymetrix Microarray Core, Gene Microarray Shared Resource, Oregon Health & Science University.

2B. CNV program evaluation:

Three programs were chosen for evaluation: dChipSNP[137], PLASQ [138] and CNAGv2 [131]. Each program was evaluated using the default settings and a common reference data set. Copy number estimation is based on a DNA change relative to a reference sample(s). Non-demented samples were used as the reference set in this study. Nine samples were chosen as the reference set due to the memory constraints of the PLASQ program. Program choice was driven by the ability of the program to determine CNV on the Affymetrix 100K mapping platform and public availability of the program. The Affymetrix CNAT 3.0 algorithm was not used in the comparison because upon initial evaluation, it did not provide for normalizing across samples. The three chosen algorithms are model based and use the information from the probes and targets across the samples to estimate copy number.

CNV results from all analyses were imported into a custom designed Microsoft Access database for storage and comparison.

2C. Gene Expression Profiling:

Differential gene expression of subjects with clinical diagnosis of AD (5 subjects) relative to non-demented subjects (9 subjects) using Affymetrix GeneChip HG-U133 Plus 2.0 arrays was performed in a previous study (Chapter II, [57]). All samples used for gene expression were also used in this study and augmented by an additional non-demented (N=25) and Probable AD (N=14) subjects. Probe sets were annotated for genomic region using Affymetrix annotation files (http://www.affymetrix.com/support/technical/byproduct.affx?product=100k). Genomic location was used to compare probe sets with the copy number variable regions in this study. Probes sets within these regions were flagged based on their differential gene expression in the whole genome expression study.

2D. Statistical Analysis and Annotation:

Statistical analyses were performed in the R v2.3.1 system for statistical computation ([126], http://www.R-project.org) using standard libraries and custom built scripts. Copy number variable regions found on the array were annotated for gene name, function, and chromosome location using NetAffx (http://www.affymetrix.com/analysis/index.affx) for PLASQ and CNAG. The dChipSNP annotation file was used for dChipSNP. All annotations were done using NCBI build 35. Annotation files were compared by Probe set ID and chromosome location. Discrepancies were evaluated using Blast ([139], NCBI).

Published CNV regions were obtained from the Database for Genomic Variants ([140], http://projects.tcag.ca/variation/) and compared to the CNV regions in this study using genomic location.

Differentially expressed genes identified in the previous study ([57], Chapter. II) were assigned to Biological Process categories of the Gene Ontology (GO) Consortium (http://www.geneontology.org/ DATE). We used GOSTAT [75] to assess representation of differentially expressed genes within CNV regions compared to all genes within CNV regions by χ^2 analysis with a FDR adjusted significance level of 0.05 [76].

Results:

3A. Concordance among Copy Number Estimation Programs is Low

Both dChipSNP and CNAG were evaluated using the Hidden Markov Model (HMM) option as this has been the most robust under the cancer tissue model [15, 58]. HMMs are statistical models in which the most likely hidden parameters are inferred from a sequence of observable parameters [141]. The copy number at a particular SNP is considered to be a “hidden” state and the process is assumed Markovian in that the probability of the copy numbers at the SNP previous and next SNPs are independent of the copy number at the present SNP. This results in a most likely “path” of copy numbers at each SNP along a chromosome. The PLASQ algorithm was evaluated using the default parameters. PLASQ is a model based algorithm that specifically models the location of the Probe set offsets and utilizes a circular binary segmentation algorithm [142] to identify segments of copy number changes along the chromosome.

There is a large discrepancy between the identified amplifications and deletions based on the choice of method (Table 1). PLASQ identified mostly deletions (11099 vs. 61 amplifications) while dChipSNP identified 83007 amplifications and 21340 deletions. CNAG identified fewer SNPs overall (2638) and all of these were included in both dChipSNP and PLASQ.

There is greater overlap between samples showing CNV with CNAG and dChipSNP than with PLASQ. Although PLASQ and dChipSNP had more CNV SNPs in common (10302), only 3703 were concordant as amplifications or deletions. Of the 2612 deletions concordant among all 3 programs, 2304 are on chromosome 23. This is due to the dosage difference of the X chromosome between males and females.

Gene annotations were based on NCBI build 35 for all three programs. Because of the dynamic nature of annotations among databases and sources, we compared the annotations provided by each program. The SNP IDs showed discrepancies for 40 genes (Table 2) between dChipSNP and Affymetrix annotations. PLASQ and CNAG annotations were identical to those of Affymetrix. Alignments of SNP probe sequences using BLAST showed homology to sequences on chromosomes listed in the dChipSNP annotations. Therefore, the dChipSNP annotations were used in all analyses.

The dChipSNP and CNAG programs were first analyzed using the HMM features available in each program because HMMs have been shown to give the accurate results. However, these algorithms were optimized for cancer phenotypes because of the high frequency of copy number changes in cancer cells [58]. Chromosomal aberrations in cancer cells generally involve large sections of highly variable DNA. It was unclear whether HMM would be applicable to this study due to the smaller predicted size of the variable regions in normal CNV. Therefore, a trimmed analysis (dChipSNP mannual) and a median smoothing method in dChipSNP were also considered. The trimmed analysis is most likely to identify rare copy number variants and not common CNV regions across the samples because it selects for the extreme values. Therefore, it was not continued because the copy number changes in normal individuals and possibly those of interest in AD may have higher frequency in the population and only the outliers are detected in a trimmed analysis. Based on the median smoothing method, there was a bias due to the date the samples were hybridized to the chips (Figure 1). Both the DM and BRLMM call rates as well as the overall chip intensity showed this dependence on hybridization date (Table 3).

CNAG corrects for both PCR fragment length and GC content [131] which the previous study determined are the primary cause of batch effects (Chapter III). Although both PLASQ and dChipSNP are model based algorithms, they were not as effective in correcting for the batch effects seen in this study. Therefore, the rest of the analyses were done using CNAG.

3B. Reference Data Set Affects Copy Number Detection:

The copy number of a particular sample is estimated relative to a reference sample. For the previous analyses, samples from the same nine non-demented subjects were used as the reference. To investigate the effect of reference set on copy number detection and estimation, several other references data sets were used. For a normal population the common options for a copy number reference are: 1) to choose one sample and compare all other samples to it, 2) choose a set of all control samples as a reference data set and compare all case samples to this data set. Because one purpose of this study is to determine the effect of copy number variation on gene expression, another option for a reference set is to use the non-demented samples from the gene expression study and compare all other samples to it. The severe batch effects necessitated the use of CNAG which has an alternate strategy for choosing a reference data set.

Data sets were examined for the purpose of reducing the batch effects. One option is to choose a fixed number of reference samples to compare to all other samples. A second option is to allow the program to chose within a given reference data set, the “best fit” of the number of samples that give the lowest SD across all SNPs. This means that the reference data set will be different for each sample. These two options were investigated with the following reference data sets:

1) the nine expression non-demented samples as the fixed set (expr all)

2) the nine expression non-demented samples as the best fit set (expr best fit

3) nine non-demented samples not part of the expression study as the fixed set (samp)

4) a set of non-demented samples with low DM call rates as the best fit set (low)

5) a set of non-demented samples with high DM call rates as the best fit set (high)

6) the total number of non-demented samples as the best fit set (total bestfit)

The reference data set using ‘options 6’, the total number of non-demented samples in the study, gave the lowest mean and SD across intensity ratios (Table 4). All other reference data sets were comparable to each other. This is probably due to both a larger sample size and allowing the CNAG algorithm to choose only those samples that match the test sample in intensity values across all SNPs. This will reduce the variability of the test sample/reference set ratio and increase the ability to identify copy number variants.

Comparison of the CNV regions as estimated by CNAG using each reference data set shows the dependence of CNV determination on reference samples. The expression samples, whether at the fixed or best fit setting, gave the highest number of CNV regions (Table 5A) or SNPs (Table 5B). The total reference data set gave the least number (CNV 526).

2C. Data Quality Affects CN Estimation:

All samples hybridized to the HindIII and XbaI chips were tested for data quality based on the following criteria: genotyping call rate, variability of $\log_2$ intensity ratios across all SNPs on the chip and the ratio of copy number variable (CNV) regions on the two chips based on the CNAG algorithm.

Differences of hybridization quality were seen between the HindIII and XbaI chips where the HindIII chip gave consistently higher call rates (Table 4). The BRLMM algorithm generally gave improved call rates over the DM algorithm for all samples. Nevertheless, there were two HindIII samples (1430B1, 1456B3) and one XbaI (1433B3) with BRLMM call rates of < 93%. These samples also had higher variability of $\log_2$ intensity ratios across all SNPs on the chip (Table sample characteristics). Standard deviations (SD) across all chips were less than 0.4 for all samples except these three.

To determine whether the variability in intensity values was sample dependent or hybridization dependent, the ratio of CNV regions of the chips relative to each other was determined (Table 6A). All ratios were between 0.18 and 6.5 except for two samples, 1430B1 (ratio HindIII: XbaI = 181), and 1433B3 (ratio HindIII: XbaI = 0.12). Sample 1456B3 had a ratio of 3.4 using only one best fit reference sample and a ratio of 57.31 using the best fit 3 reference samples. A ratio different from 1 shows an imbalance in hybridizations between chips and suggests technical difficulties. These three samples (1430B1, 1433B3, 1456B3) were removed from the final data set.

Removing these problematic samples from the final data set reduced the variability across all SNPs (Table 7). The reduced data set shows a decrease in CNV regions and CNV SNPs (Table 8). Contiguous CNV SNPs are considered one CNV region. The concordance of CNV regions and CNV SNPs between the data sets using all control samples as references (total refs) and the data set without the three samples (reduced set) showed that 97% of the reduced set is concordant with the full data set (Table 8).

The reduced data set, therefore, was chosen as the final data set. Analyses of CNV on chromosome 23 was done separately using the reduced data set and only same sex references were used to determine copy number on the X chromosome (Table 7).

Finally, all SNPs that had an intensity $\log_2$ ratio = 0 across all samples were removed from further analysis (12 SNPs).

Throughout the analyses, the HindIII and XbaI chips were analyzed separately. The rationale for this was twofold. First, the pattern of variability across the two chips was different as shown by the genotyping call rates and the SD across all SNPs. Secondly, the control references used to determine CNV were different for each sample and for each chip due to the criteria of fitting only those sample references that did not increase the SD across the SNPs. Therefore, CNV regions on each chip are relative to only those samples used as the references and may be different for the two chips. The number of CNV per sample varies from 0 to 6 across all samples (data not shown).

3D. Distribution of CNV regions:

A total of 295 regions across the genome were found to be copy number variable (Table 9A). All chromosomes contained regions of CNV. All samples in the study except for one had CNV regions (Table 10). Of these, 268 showed a gain in DNA, 19 a loss and 8 regions had samples with either a loss or a gain. All chromosomes showed multiple CNV regions with the larger chromosomes containing more regions.

Chromosome 4 had the highest number of CNV regions (26), whereas chromosome 22 had the least (3). None of the discrepant SNPs (Table 2) were located in CNV regions.

CNV regions were identified based on 1- 36 SNPs with the largest region being on chromosome 23. Multiple samples with copy number variants were found in 29 regions. Twelve of these regions have been previously published. Regions on chromosomes 14 and 15 were the most variable across the samples in this study (10 and 26 samples, respectively).

Genomic regions previously associated with AD through genetic mapping studies were compared to CNV regions (Table 11). All AD linked regions (87.5%) except 1q23-31 and 9p21 contain CNV regions. Twelve of these CNV regions are supported by more than one SNP and 8 CNV regions in 3 AD linked regions are represented by more than one sample.

3E. Comparison with CNV regions previously published:

Previously published CNV regions were downloaded from the Database of Genomic Variants (http://projects.tcag.ca/variation/). A total of 2191 loci are known to be CNV. Of the 295 regions found in this study, 68 have been identified previously (Tables 9B and 12). Five regions are located within AD linked regions (Table 11).

CNV regions supported by more than one sample found in this study but not in previous studies were located in 19 regions on 11 chromosomes (Tables 10 and 12). Six of these regions were supported by more than one SNP.

The reference data set defines the copy number change in a data set. Therefore, copy number variants may show gains or losses differently from study to study. This has implications for a meta analysis and suggests that a literature based approach may not be as informative for absolute CN variability estimates.

3F. Comparison of CNV regions with differentially expressed probe sets:

Previously studied gene expression differences in AD subjects relative to controls were used to determine the possible effect of CNV regions on gene expression results. All probe sets on the HG133 Plus2 array were combined with the CNV regions found in this study (Table 13). Of the 54,000 probes sets on the array, 855 were located in the CNV regions. Of these, 119 probe sets were differentially expressed and 112 were located within the coding regions of the genes (Table 13). Seventy-one were located in previously published CNV regions. Multiple samples were variable for 38 probe sets located in 7 CNV regions.

Analysis for over or under representation of Biological Processes in the GO categories for the differentially expressed genes relative to all genes in the CNV regions showed no processes significantly over or under represented than expected by chance.

4. Discussion:

Recent discoveries of the extent of copy number variants (CNVs) across the human genome indicate that normal genetic variation encompasses a wider range of genomic architecture than previously thought and raises the question of whether these variants could influence complex traits. To determine the impact of CNVs on gene expression, we compared copy number variation and gene expression in a population of cognitively healthy individuals and individuals with Alzheimer's disease (AD). We identified 280 loci across the genome that are copy number variable in our population. Nineteen of these regions were unique to this study. Comparison with differentially expressed genes revealed that 119 differentially expressed probe sets were located within copy number variable regions. Overall, 22.9% of differentially expressed genes in our population could be affected by copy number variability.

4.1 Copy number algorithm affects identification of CNV

The significance of the impact due to the underlying model used to determine the CNV regions is highlighted by the low concordance of copy number variable regions identified by the different algorithms. Of the three algorithms tested in this study, only CNAG corrects for the PCR fragment length and the GC content of the probes. This is crucial for reducing the variability in the data and obtaining accurate relative copy numbers. The option available in the CNAG algorithm that compares all reference samples to each test sample and selects the least variable reference sample set based on standard deviation across all the samples. This is a stringent method of CNV detection which should allow for more false negatives while controlling for fewer false positives. Chips processed in the same batch had lower standard deviation than those across batches. Thus, the reference samples selected tended to be samples hybridized in the same batch as the test samples. For an unpaired experimental design, batch effects had the greatest impact on relative intensity values between test and reference samples. Copy number algorithms correcting for batch effects are more accurate in identifying copy number differences.

CNV regions comprising single SNPs are of lower confidence than CNV regions with more than one SNP. Nevertheless, these SNPs were kept in the analysis as CNVs <100kb in length are more numerous in the genome than CNVs >100kb [13]. Studies using arrays of higher SNP density would provide higher resolution to refine the regions of CNV.

4.2 CNV regions impact differentially expressed genes

Gene regulation is a complex process and CNV is one possible mechanism for altering expression of genes located within the CNV region and of genes in downstream cascades. This study shows that differentially expressed genes are located within CNV regions and that nearly all samples had CNV regions impacting these genes. Many of the CNVs were found in one or a few individuals suggesting substantial individual differences in gene copy number across the genome [8, 143]. The most common genes affected by CNVs were the olfactory receptors and the immunoglobulin genes as has been seen in previous studies [8]. Overall, 22.9% of the differentially expressed probe sets were possibly affected by CNV suggesting a role for CNV in functionally relevant genetic variation. Expression within a CNV region was not necessarily affected by the number of genes suggesting that gene expression regulation networks can absorb much of the “normal” copy number variation. Nevertheless, these results indicate the need for considering the possibility of an underlying copy number difference in the altered expression in gene profiling studies.

4.3 Considerations for Copy Number Determination in non-cancer tissues

As stated earlier, the algorithms used to determine copy number differences in this study were designed and optimized for cancer tissues which have a substantially different type of structural variation and genomic distribution from normal cells. All subjects in this study would have CNVs characteristic of normal individuals as any structural variation characteristic of AD would still be more focused and of smaller magnitude than those seen in cancer cells. Smaller, more focused segments of CNV and fewer copy number extremes present a challenge for the current algorithms. Major considerations were to explore appropriate reference data sets and to accommodate for batch effects within a study. Intensity differences across batches increase the $\log_2$ ratio variability and decrease ability to identify copy number changes. CNAG adjusts for batch effects by correcting for the PCR fragment length and probe GC content. Improvements in modeling the underlying process as well as increased SNP density are needed to increase the ability to identify copy number variants.

We evaluated the impact of CNV on gene expression within the limitations of our sample size, SNP coverage and CNV resolution. Caveats about these data are as follows:

1) there will be some CNVs not identified because the reference data set changes based on minimum standard deviation

2) some CNVs could be artifacts based on the number of SNPs used to call a CNV

3) the resolution of the chip to identify CNVs is dependent on SNP density and so will vary across the genome

4) likewise, boundaries of CNVs are dependent on local SNP density and therefore, CNVs smaller than interSNP density will be missed

5) inversions and balanced rearrangements will not be detected

As seen in other studies [8], the majority of CNVs are gains, probably owing to a greater tolerance in the genome for large gains versus deletions. This also probably reflects the difficulty of detecting deletions due to the greater variability in lower intensity ranges.

Although it lacks the power to investigate CNV effects on phenotype, this pilot study suggests that CNV effects may be individual and gene specific. This has been seen in other studies where an increase in CN variability is seen in a trait, but the specific region affected varies among individuals [144, 145].

Chapter IV. Tables.

Table 1. Concordance of SNPs among CN programs

A. Total number of SNPs concordant among the 3 methods

	PLASQ	dchip	cnag
PLASQ	11160	10302	2638
dchip		104347	2638
cnag			2638

PLASQ/dchip/cnag = 2638

Table 2. Discrepancies between the Affymterix and dChipSNP annotations

Affymetrix			dChipSNP		
probeID	**chr**	**pos**	**probeID**	**chr**	**pos**
SNP_A-1648117	1	37194096	SNP_A-1648117	8	37132222
SNP_A-1678541	2	7599411	SNP_A-1678541	5	7599149
SNP_A-1725297	2	25127948	SNP_A-1725297	3	25127909
SNP_A-1756770	2	35027805	SNP_A-1756770	3	35024569
SNP_A-1695177	2	39527680	SNP_A-1695177	3	39513268
SNP_A-1681558	2	108947883	SNP_A-1681558	4	109049099
SNP_A-1707690	3	18595288	SNP_A-1707690	11	18603021
SNP_A-1643381	3	22132614	SNP_A-1643381	20	22179614
SNP_A-1720364	3	39561067	SNP_A-1720364	12	39561067
SNP_A-1708353	3	73985190	SNP_A-1708353	8	73872597
SNP_A-1753744	3	86003512	SNP_A-1753744	14	84923800
SNP_A-1723789	3	116417029	SNP_A-1723789	6	116355906
SNP_A-1704638	4	26882729	SNP_A-1704638	16	26941778
SNP_A-1686768	4	52935752	SNP_A-1686768	14	51856040
SNP_A-1754495	4	54567036	SNP_A-1754495	8	54454443
SNP_A-1748811	4	104367631	SNP_A-1748811	13	103267631
SNP_A-1717206	4	136045098	SNP_A-1717206	7	136003885
SNP_A-1705265	4	136382126	SNP_A-1705265	5	136430443
SNP_A-1753755	5	5453387	SNP_A-1753755	11	5461120
SNP_A-1755468	5	24958546	SNP_A-1755468	14	23878834
SNP_A-1710238	5	80558774	SNP_A-1710238	13	79458774
SNP_A-1658231	6	66419538	SNP_A-1658231	10	66094135
SNP_A-1690305	6	118221091	SNP_A-1690305	10	117895688
SNP_A-1671975	7	19682770	SNP_A-1671975	16	19741822
SNP_A-1674753	7	33916496	SNP_A-1674753	9	33916496
SNP_A-1756363	7	83012428	SNP_A-1756363	11	83061088
SNP_A-1704345	8	405401	SNP_A-1704345	18	405401
SNP_A-1741271	8	62090667	SNP_A-1741271	16	63308379
SNP_A-1682585	9	113115250	SNP_A-1682585	7	128603353
SNP_A-1744215	9	130697593	SNP_A-1744215	11	130730035
SNP_A-1647874	10	25804577	SNP_A-1647874	21	25804577
SNP_A-1667621	10	64877754	SNP_A-1667621	12	64877754
SNP_A-1664929	10	97536416	SNP_A-1664929	13	96436416
SNP_A-1645865	11	58376174	SNP_A-1645865	18	58374163

SNP_A-1738821	11	81581498	SNP_A-1738821	13	80481498
SNP_A-1737106	12	21619612	SNP_A-1737106	21	21619612
SNP_A-1670271	12	41776810	SNP_A-1670271	15	41705574
SNP_A-1708602	12	41941636	SNP_A-1708602	14	40861924
SNP_A-1665266	13	38936088	SNP_A-1665266	23	38210772
SNP_A-1658886	22	45975243	SNP_A-1658886	23	45264971

Table 3. Affect of Hybridization Date on log$_2$ratio data [a]

		Expr					Total
chip	**intensity**	**bestfit[b]**	**expall**	**low**	**high**	**samp**	**bestfit**
XbaI	8.55E-05	2.68E-07	2.81E-07	2.89E-06	1.56E-06	6.51E-05	0.0006809
HindIII	0.06601	4.99E-08	4.99E-08	1.36E-06	1.28E-07	1.73E-06	3.33E-06

[a] p-values from the Kruskal-Wallis rank sum test. Only dates on which >2 samples were hybridized wer e used.

[b] reference data sets are named as follows:

1) the nine expression non-demented samples as the fixed set (expr all)
2) the nine expression non-demented samples as the best fit set (expr best fit
3) nine non-demented samples not part of the expression study as the fixed set (samp)
4) a set of non-demented samples with low DM call rates as the best fit set (low)
5) a set of non-demented samples with high DM call rates as the best fit set (high)
6) the total number of non-demented samples as the best fit set (total bestfit)

Table 4. Summary statistics for array data across all samples[a]

HindIII

	DM	BRLMM		Expr					Total
	call	**call**	**intensity**	**bestfit.SD**	**expall.SD**	**samp.SD**	**low.SD**	**high.SD**	**refs.SD**
Mean	97.656	99.226	10.954	0.253	0.253	0.251	0.259	0.262	0.172
SD	2.175	1.378	0.289	0.086	0.086	0.080	0.091	0.086	0.061
Max	99.640	99.920	11.571	0.524	0.524	0.607	0.692	0.568	0.514
Min	89.210	92.890	9.843	0.145	0.145	0.174	0.154	0.159	0.112

XbaI

	DM	BRLMM		Expr					Total
	call	**call**	**intensity**	**bestfit.SD**	**expall.SD**	**samp.SD**	**low.SD**	**high.SD**	**refs.SD**
Mean	94.116	98.540	10.688	0.259	0.259	0.240	0.247	0.254	0.169
SD	2.811	1.028	0.303	0.070	0.069	0.044	0.050	0.059	0.035
Max	99.270	99.670	11.323	0.471	0.471	0.435	0.394	0.462	0.323
Min	82.920	92.810	9.975	0.143	0.143	0.154	0.152	0.162	0.113

[a] reference data sets are named as follows:

1) the nine expression non-demented samples as the fixed set (expr all)
2) the nine expression non-demented samples as the best fit set (expr best fit
3) nine non-demented samples not part of the expression study as the fixed set (samp)
4) a set of non-demented samples with low DM call rates as the best fit set (low)
5) a set of non-demented samples with high DM call rates as the best fit set (high)
6) the total number of non-demented samples as the best fit set (total bestfit)

Table 5. Copy number variable regions based on CNAG analysis using different reference sample sets [a]

A. regions	expr bestfit[b]	expr 9 fixed	Low DM call rate	High DM call rate	9 fixed samples	Total refs bestfit
expr bestfit	1076	726	301	564	379	310
expr 9 fixed		1038	322	626	403	279
Low DM call rate			605	335	385	263
High DM call rate				719	442	267
9 fixed samples					593	269
Total refs bestfit						526

B. SNPs	expr bestfit	expr 9 fixed	Low DM call rate	High DM call rate	9 fixed samples	Total refs bestfit
expr bestfit	95 (12073)	83 (9575)	57 (5688)	61 (8139)	55 (6108)	49 (5568)
expr 9 fixed		110 (11953)	80 (5981)	85 (8925)	79 (6500)	61 (4742)
Low DM call rate			247 (10141)	85 (5986)	131 (6209)	95 (4084)
High DM call rate				95 (10100)	84 (7165)	66 (4414)
9 fixed samples					179 (8510)	86 (3911)
Total refs bestfit						119 (5935)

[a] autosomes only

[b] reference data sets are named as follows:

1) the nine expression non-demented samples as the fixed set (expr 9 fixed)
2) the nine expression non-demented samples as the best fit set (expr bestfit
3) nine non-demented samples not part of the expression study as the fixed set (9 fixed samples)
4) a set of non-demented samples with low DM call rates as the best fit set (low DM call rate)
5) a set of non-demented samples with high DM call rates as the best fit set (high DN call rate)
6) the total number of non-demented samples as the best fit set (total refs bestfit)

Table 6. Copy numbers for HindIII and XbaI chips [a]

	A. total best fit reference data set[b]			**B. reduced total reference set**		
sample	**hind**	**xba**	**ratio**	**hind**	**xba**	**ratio**
X464T3	3	15	0.2	3	15	0.200
X497B3	112	17	6.5882353	112	17	6.588
X515B3	6	10	0.6	6	10	0.600
X555B3	7	10	0.7	7	10	0.700
X588T1	42	12	3.5	42	12	3.500
X597B3	44	30	1.4666667	44	30	1.467
X615B3	23	15	1.5333333	23	15	1.533
X538B3	8	15	0.5333333	8	15	0.533
X840B1	7	22	0.3181818	7	22	0.318

X946B3	8	10	0.8	8	10	0.800
X956T1	64	21	3.047619	64	21	3.048
X1009B3	8	27	0.2962963	8	27	0.296
X1081B2	2	10	0.2	2	10	0.200
X1116B1	25	19	1.3157895	25	19	1.316
X1430B1	5263	29	181.48276	0	29	0.000
X1456B3	34	10	3.4	0	10	0.000
X1476T3	2	11	0.1818182	2	11	0.182
X1542T1	10	13	0.7692308	10	13	0.769
X783B3	12	10	1.2	12	10	1.200
X848B1	2	10	0.2	2	10	0.200
X958B1	11	10	1.1	11	10	1.100
X993B3	14	16	0.875	14	16	0.875
X1013B1	14	25	0.56	11	25	0.440
X1052T1	11	34	0.3235294	11	34	0.324
X1065B3	3	10	0.3	3	10	0.300
X1104B2	3	34	0.0882353	3	34	0.088
X1153B3	3	34	0.0882353	3	34	0.088
X1166B3	8	13	0.6153846	8	13	0.615
X1317T2	7	18	0.3888889	7	18	0.389
X1395B3	5	10	0.5	5	10	0.500
X1399B3	9	22	0.4090909	9	22	0.409
X1401B3	13	36	0.3611111	13	36	0.361
X1432B1	2	10	0.2	2	10	0.200
X1433B3	14	117	0.1196581	14	0	Inf
X1466T2	10	14	0.7142857	13	14	0.929
X1469B1	7	27	0.2592593	7	27	0.259
X1488T3	6	23	0.2608696	6	23	0.261
X810T1	14	14	1	14	14	1.000
X869T1	9	29	0.3103448	16	27	0.593
X334T2	11	36	0.3055556	11	36	0.306
X1014T2	16	31	0.516129	16	31	0.516
X1062T1	2	15	0.1333333	2	15	0.133
X1229T2	7	29	0.2413793	7	29	0.241
X1376T2	13	15	0.8666667	8	15	0.533
X719T3	10	13	0.7692308	10	13	0.769
X1309T1	14	11	1.2727273	14	11	1.273
X862T1	7	24	0.2916667	6	24	0.250
X1538T3	7	10	0.7	7	10	0.700
X1524T2	7	11	0.6363636	7	11	0.636
X1547T2	9	32	0.28125	9	32	0.281
X486T2	10	15	0.6666667	17	15	1.133
X697T2a	11	24	0.4583333	8	26	0.308
X699T2a	8	10	0.8	9	11	0.818
X1390T3a	10	28	0.3571429	10	28	0.357

[a] Autosomes only. Copy numbers other than two.

[b] total best fit reference data set compared to the total best fit set with the three low quality sampels removed.

Table 7. Effect of data quality on variability across SNPs.

	HindIII			XbaI		
	total.sd[a]	CNVshort	CNV23	total.sd	CNVshort	CNV23
Mean	0.172	0.163	0.177	0.169	0.165	0.179
SD	0.061	0.036	0.038	0.035	0.028	0.037
Median	0.156	0.155	0.166	0.163	0.163	0.175
Min	0.112	0.112	0.112	0.113	0.113	0.113
Max	0.514	0.263	0.283	0.323	0.233	0.279

[a] total best fit reference data set compared to the total best fit set with the three low quality sampels removed.

Table 8. Concordance of autosomal CNV between reference data sets

	regions		SNPs	
	Total refs bestfit [a]	reduced set [b]	Total refs bestfit	reduced set
Total refs bestfit	45(481) [c]	41(311)	119(5935)	101(607)
reduced set		42(321)		107(613)

[a] full data set

[b] reference data set with 3 low qualitly samples removed

[c] deletions (amplifications)

Table 9. Pattern of CNV regions and SNPs across chromosomes.

A. Number of CN differences in this study

	CNV regions			CNV SNPs	
chr	loss	gain	loss/gain	loss	gain
1	0	12	0	0	16
2	1	18	0	3	77
3	3	16	0	13	44
4	2	26	0	3	46
5	2	21	0	5	34
6	0	23	0	0	53
7	0	21	0	0	69
8	2	15	1	18	22
9	1	6	1	3	20
10	1	12	2	5	28
11	1	16	0	1	37
12	1	8	0	3	10
13	1	7	0	2	10
14	2	14	1	28	28
15	0	6	2	0	58
16	0	4	0	0	10
17	0	5	0	0	10
18	0	3	0	0	5
19	0	6	0	0	10
20	0	5	0	0	17
21	2	4	0	11	6
22	0	3	0	0	3
23	0	17	1	1	222

B. Known CNV regions[a]

	# regions concordant		
chr	loss	gain	loss/gain
1	0	2	0
2	0	4	0
3	2	4	0
4	0	1	0
5	0	3	0
6	0	5	0
7	0	2	0
8	2	4	1
9	0	1	1
10	0	4	1
11	1	2	0
12	0	2	0
13	0	0	0
14	1	7	1
15	0	1	2
16	0	1	0
17	0	3	0
18	0	0	0
19	0	0	0
20	0	1	0
21	0	1	0
22	0	2	0
23	0	6	0

[a] Known CNV regions were obtained from Database for Genomic Variants ([137], http://projects.tcag.ca/variation/) and compared with regions found in this study.

Table 10. CNV regions for all samples[a]

sample	chip	CN[b]	chr	cytoband	StartSNPc	StartPos	EndSNP	EndPos	linkedd	more than 1 SNP[e]	more than 1 sample[f]	region[g]	# samples[h]	Published region[i]
588T1	hind	5	1	p36.22	65	9898974	66	9924378	Y	1	0	1	1	0
597B3	hind	6	1	p36.12	183	21121210	183	21121210	Y	0	0	2	1	0
697T2a	xba	5	1	p35.1	57539	34168985	57539	34168985	N	0	0	3	1	0
1390	xba	5	1	p34.3	57603	38018699	57603	38018699	N	0	0	4	1	0
588T1	hind	6	1	p32.3	489	50766343	489	50766343	N	0	0	5	1	0
588T1	hind	5	1	p32.1	705	59378042	706	59436588	N	1	0	6	1	0
597B3	hind	6	1	p31.3	829	64332630	829	64332630	N	0	0	7	1	0
1547T2	hind	6	1	p31.1	1057	72104806	1057	72104806	N	0	0	8	1	0
1013B1	xba	5	1	p22.1	58999	94031284	58999	94031284	N	0	0	9	1	0
1488T3	xba	5	1	q25.3	60452	177707522	60453	177707748	N	1	0	10	1	0
956T1	hind	6	1	q41	4062	218728033	4062	218728033	N	0	0	11	1	1
1466T2	hind	4	1	q44	4521	242779667	4522	242780409	N	1	0	12	1	1
588T1	hind	0	2	p24.3	4751	12255441	4753	12255820	Y	1	0	13	1	0
1052T1	xba	6	2	p21	63000	45253577	63000	45253577	N	0	0	14	1	0
810T1	xba	6	2	p16.3	63056	47985466	63056	47985466	N	0	0	15	1	0
1430B1	xba	4	2	p13.3	63721	70796231	63725	70975597	N	1	0	16	1	0
956T1	xba	6	2	p12	63803	76475818	63803	76475818	N	0	0	17	1	1
1390	xba	5	2	q11.1	64104	95606297	64104	95606297	N	0	0	18	1	1
1430B1	xba	5	2	q11.2	64150	101375977	64152	101495148	N	1	0	19	1	0
497B3	hind	3	2	q14.1	6798	116629374	6824	117245131	N	1	0	20	1	1
597B3	hind	6	2	q21.3	7132	136300896	7132	136300896	N	0	0	21	1	0
1104B2	xba	3	2	q21.3	64760	137094155	64783	137459907	N	1	0	22	1	0
1052T1	xba	6	2	q31.1	65751	173183764	65751	173183764	N	0	0	23	1	0
486T2	hind	5	2	q31.2	8269	179038313	8270	179038998	N	1	0	24	1	0
597B3	hind	6	2	q32.2	8559	190829101	8559	190829101	N	0	0	25	1	0
810T1	hind	6	2	q36.2	9400	225729560	9400	225729560	N	0	1	26	3	0
334T2	hind	6	2	q36.2	9400	225729560	9400	225729560	N	0	1	26	3	0
1547T2	hind	6	2	q36.2	9400	225729560	9400	225729560	N	0	1	26	3	0
956T1	hind	6	2	q36.3	9519	230126120	9519	230126120	N	0	0	27	1	0
1013B1	hind	5	2	q37.1	9578	235240125	9578	235240125	N	0	0	28	1	0
1166B3	hind	3	2	q37.3	9604	241800408	9609	242399567	N	1	1	29	3	1
1433B3	hind	3	2	q37.3	9604	241800408	9609	242399567	N	1	1	29	3	1
1052T1	hind	5	2	q37.3	9606	241802208	9606	241802208	N	0	1	29	3	0
1052T1	hind	3	2	q37.3	9607	242141304	9609	242399567	N	1	0	30	1	0
615B3	hind	3	3	p26.3	9610	48603	9623	186745	N	1	0	31	1	1
1309T1	hind	0	3	p26.3	9689	1510440	9690	1511278	N	1	0	32	1	1

334T2	xba	3	3	p26.3	67210	1643488	67222	1686749	N	1	0	33	1	0
783B3	hind	1	3	p26.1	9838	4110776	9847	4136335	N	1	0	34	1	1
956T1	hind	5	3	p21.31	10748	44828819	10750	44858975	N	1	0	35	1	0
1013B1	xba	0	3	p21.31	68245	45192647	68245	45192647	N	0	0	36	1	0
597B3	hind	5	3	p14.3	10845	55348944	10845	55348944	N	0	0	37	1	0
1014T2	hind	6	3	p14.3	10868	56656693	10868	56656693	N	0	0	38	1	0
597B3	hind	6	3	p14.2	10939	60189407	10939	60189407	N	0	0	39	1	0
956T1	hind	6	3	p14.1	11178	66795667	11178	66795667	N	0	0	40	1	0
956T1	hind	6	3	p13	11359	73989706	11359	73989706	N	0	0	41	1	0
956T1	xba	6	3	q11.2	69108	97317545	69108	97317545	N	0	0	42	1	0
1430B1	xba	6	3	q12.2	69181	102254723	69181	102254723	N	0	0	43	1	1
597B3	hind	6	3	q13.31	12073	116705694	12073	116705694	N	0	0	44	1	0
1052T1	xba	6	3	q13.33	69627	121161492	69627	121161492	N	0	0	45	1	0
1466T2	xba	5	3	q23	69972	143280552	69972	143280552	N	0	0	46	1	1
1052T1	xba	6	3	q25.1	70186	152942205	70186	152942205	N	0	0	47	1	0
1166B3	xba	5	3	q27.3	70866	188644318	70866	188644318	N	0	1	48	2	0
1309T1	xba	6	3	q27.3	70866	188644318	70866	188644318	N	0	1	48	2	0
588T1	hind	5	3	q29	13560	197963623	13561	198449496	N	1	0	49	1	1
1013B1	xba	4	4	p15.2	71530	26115291	71533	26115900	N	1	0	50	1	0
1014T2	xba	6	4	p14	71787	37005212	71787	37005212	N	0	0	51	1	0
538B3	hind	6	4	q12	14746	59689945	14746	59689945	N	0	1	52	2	0
1469B1	hind	6	4	q12	14746	59689945	14746	59689945	N	0	1	52	2	0
588T1	hind	6	4	q13.1	14772	60870592	14772	60870592	N	0	0	53	1	0
1052T1	xba	4	4	q13.1	72293	62566178	72294	62566224	N	1	0	54	1	0
597B3	hind	6	4	q13.2	15006	68577813	15006	68577813	N	0	0	55	1	0
334T2	hind	0	4	q21.22	15371	83208150	15371	83208150	N	0	0	56	1	0
334T2	hind	4	4	q21.22	15372	83239664	15374	83246577	N	1	0	57	1	0
464T3	hind	5	4	q21.23	15436	86201336	15436	86201336	N	0	1	58	4	0
1488T3	hind	5	4	q21.23	15436	86201336	15436	86201336	N	0	1	58	4	0
334T2	hind	5	4	q21.23	15436	86201336	15436	86201336	N	0	1	58	4	0
699T2a	hind	5	4	q21.23	15436	86201336	15436	86201336	N	0	1	58	4	0
1052T1	hind	6	4	q22.1	15614	91598041	15614	91598041	N	0	0	59	1	0
1014T2	xba	6	4	q22.3	73068	94178695	73068	94178695	N	0	0	60	1	0
697T2a	xba	4	4	q22.3	73160	97585566	73162	97834662	N	1	0	61	1	0
588T1	hind	6	4	q24	15984	107872099	15984	107872099	N	0	0	62	1	0
1166B3	xba	5	4	q25	73515	112835774	73515	112835774	N	0	0	63	1	0
588T1	xba	6	4	q26	73632	120076082	73632	120076082	N	0	0	64	1	0
956T1	hind	6	4	q27	16315	121928243	16315	121928243	N	0	0	65	1	0
1153B3	xba	4	4	q27	73691	121997169	73693	122018813	N	1	0	66	1	0
1065B3	hind	5	4	q28.1	16373	124471655	16373	124471655	N	0	0	67	1	0
869T1	hind	5	4	q28.1	16425	126736617	16427	126755739	N	1	0	68	1	0
1014T2	xba	5	4	q28.2	73866	129417818	73866	129417818	N	0	0	69	1	0
588T1	xba	6	4	q28.2	73891	130916643	73891	130916643	N	0	0	70	1	0

699T2a	hind	0	4	q28.3	16546	131897592	16547	131898171	N	1	0	71	1	0
1433B3	hind	5	4	q28.3	16690	139002535	16690	139002535	N	0	0	72	1	1
1062T1	xba	6	4	q31.21	74171	143600755	74171	143600755	N	0	0	73	1	0
588T1	hind	5	4	q31.3	17046	155757047	17048	155766635	N	1	0	74	1	0
958B1	hind	3	4	q32.2	17224	162311318	17230	162359542	N	1	0	75	1	0
464T3	xba	6	4	q34.1	74872	173140931	74872	173140931	N	0	1	76	4	0
1476T3	xba	6	4	q34.1	74872	173140931	74872	173140931	N	0	1	76	4	0
862T1	xba	6	4	q34.1	74872	173140931	74872	173140931	N	0	1	76	4	0
699T2a	xba	6	4	q34.1	74872	173140931	74872	173140931	N	0	1	76	4	0
588T1	hind	6	4	q35.2	17888	188944515	17888	188944515	Y	0	0	77	1	0
1488T3	xba	6	5	p15.2	75461	9930169	75461	9930169	Y	0	0	78	1	1
1014T2	hind	4	5	p14.1	18567	28803744	18569	28947927	N	1	0	79	1	0
1430B1	xba	5	5	q11.2	76352	54289338	76353	54326926	N	1	0	80	1	0
1052T1	xba	5	5	q11.2	76407	57087315	76407	57087315	N	0	0	81	1	0
1547T2	hind	6	5	q12.1	19286	60508810	19286	60508810	N	0	0	82	1	0
1052T1	xba	6	5	q13.1	76668	67696704	76668	67696704	N	0	0	83	1	0
946B3	hind	4	5	q13.2	19555	72077661	19559	72216133	N	1	0	84	1	0
597B3	hind	4	5	q14.1	19723	79541752	19724	79780662	N	1	1	85	2	0
1013B1	hind	6	5	q14.1	19724	79780662	19724	79780662	N	0	1	85	2	0
597B3	hind	6	5	q14.1	19725	79797970	19725	79797970	N	0	0	86	1	0
699T2a	hind	6	5	q14.3	19840	84010157	19840	84010157	N	0	0	87	1	0
1014T2	xba	5	5	q14.3	77030	85380111	77032	85464871	N	1	0	88	1	1
597B3	hind	5	5	q14.3	20005	91161653	20005	91161653	N	0	0	89	1	0
862T1	xba	0	5	q21.3	77626	109452450	77628	109453291	N	1	0	90	1	0
1390	xba	6	5	q22.1	77702	111453254	77702	111453254	N	0	0	91	1	0
1309T1	hind	0	5	q22.2	20487	111710742	20488	111711216	N	1	0	92	1	0
1062T1	xba	6	5	q23.2	78132	125467810	78132	125467810	N	0	0	93	1	0
956T1	hind	6	5	q23.3	21011	129544898	21011	129544898	N	0	0	94	1	0
588T1	hind	6	5	q31.3	21194	140537521	21194	140537521	N	0	0	95	1	1
464T3	xba	5	5	q32	78549	145019024	78551	145019413	N	1	1	96	2	0
1376T2	xba	4	5	q32	78549	145019024	78551	145019413	N	1	1	96	2	0
1014T2	xba	5	5	q32	78631	147855944	78632	147857643	N	1	0	97	1	0
956T1	hind	6	5	q34	21769	162599727	21770	162599993	N	1	0	98	1	0
1547T2	hind	6	5	q35.1	21997	170695382	21997	170695382	N	0	0	99	1	0
956T1	hind	6	6	p25.3	22173	2204327	22173	2204327	N	0	0	100	1	0
597B3	hind	6	6	p22.2	22779	24753174	22779	24753174	N	0	0	101	1	0
862T1	hind	5	6	p21.31	22942	36118128	22942	36118128	Y	0	1	102	2	0
697T2a	hind	5	6	p21.31	22942	36118128	22942	36118128	Y	0	1	102	2	0
497B3	hind	3	6	p21.1	23026	41376510	23040	42959475	Y	1	0	103	1	1
1052T1	xba	6	6	q12	80862	65739442	80862	65739442	N	0	0	104	1	0
810T1	xba	6	6	q12	80894	66572761	80894	66572761	N	0	0	105	1	0
597B3	hind	6	6	q13	23643	70297715	23643	70297715	N	0	0	106	1	0
597B3	xba	5	6	q13	81088	72646417	81090	72783765	N	1	0	107	1	0

1390	xba	6	6	q14.1	81221	79016785	81221	79016785	N	0	0	108	1	1
1466T2	xba	6	6	q14.3	81436	86011003	81436	86011003	N	0	0	109	1	0
719T3	xba	4	6	q16.1	81617	93552796	81619	93553574	Y	1	0	110	1	0
956T1	hind	6	6	q16.3	24474	101905980	24474	101905980	Y	0	0	111	1	0
1014T2	hind	4	6	q22.1	24753	115417501	24756	115420450	N	1	0	112	1	1
1430B1	xba	6	6	q22.1	82191	116911322	82191	116911322	N	0	0	113	1	0
1430B1	xba	5	6	q22.31	82354	122735532	82357	122848520	N	1	0	114	1	1
588T1	hind	6	6	q22.31	25043	125539447	25043	125539447	N	0	0	115	1	0
956T1	hind	5	6	q23.2	25266	132725748	25269	132769586	N	1	0	116	1	0
956T1	hind	6	6	q23.3	25328	135385192	25328	135385192	N	0	0	117	1	0
956T1	hind	6	6	q24.1	25417	139543034	25417	139543034	N	0	0	118	1	0
956T1	hind	5	6	q24.2	25514	143850706	25516	143917617	N	1	0	119	1	1
956T1	hind	6	6	q25.1	25669	150701028	25669	150701028	Y	0	0	120	1	0
1309T1	hind	5	6	q25.1	25670	150772388	25671	150772680	Y	1	0	121	1	0
956T1	hind	6	6	q25.1	25690	151791227	25690	151791227	Y	0	0	122	1	0
956T1	hind	6	7	p21.3	26322	10458942	26322	10458942	N	0	0	123	1	0
719T3	hind	6	7	p15.3	26797	21469329	26797	21469329	N	0	1	124	2	0
862T1	hind	6	7	p15.3	26797	21469329	26797	21469329	N	0	1	124	2	0
1062T1	xba	5	7	p15.1	84367	28138715	84367	28138715	N	0	0	125	1	0
697T2a	xba	5	7	p14.1	84589	37000570	84590	37000595	N	1	0	126	1	0
956T1	hind	4	7	p13	27387	43309005	27392	43661716	N	1	0	127	1	0
597B3	hind	6	7	p11.2	27606	54184265	27606	54184265	N	0	0	128	1	0
1390	hind	5	7	p11.2	27643	56471248	27645	56479844	N	1	0	129	1	0
993B3	hind	3	7	q11.21	27653	61534066	27659	62122172	N	1	0	130	1	1
497B3	hind	3	7	q11.21	27674	65551174	27685	67019849	N	1	0	131	1	1
597B3	hind	6	7	q11.22	27752	70966450	27752	70966450	N	0	0	132	1	0
956T1	hind	6	7	q21.11	27867	79185389	27867	79185389	N	0	0	133	1	0
497B3	hind	3	7	q21.11	28053	85064707	28071	85771585	N	1	0	134	1	0
993B3	xba	6	7	q21.13	85633	89764266	85633	89764266	N	0	0	135	1	0
597B3	hind	6	7	q21.3	28305	94919442	28305	94919442	N	0	0	136	1	0
956T1	hind	5	7	q22.2	28439	103610597	28441	103762352	N	1	0	137	1	0
588T1	hind	5	7	q31.1	28605	109786112	28606	109819795	N	1	0	138	1	0
515B3	hind	4	7	q31.32	28919	122781982	28921	123001226	N	1	0	139	1	0
1390	xba	6	7	q32.1	86465	126052305	86465	126052305	N	0	0	140	1	0
956T1	xba	5	7	q32.1	86518	128719303	86518	128719303	N	0	0	141	1	0
1390	xba	6	7	q33	86723	136663169	86723	136663169	N	0	0	142	1	0
946B3	hind	6	7	q35	29408	146996311	29408	146996311	N	0	1	143	3	0
1009B3	hind	6	7	q35	29408	146996311	29408	146996311	N	0	1	143	3	0
1116B1	hind	6	7	q35	29408	146996311	29408	146996311	N	0	1	143	3	0
588T1	hind	6	8	p23.3	29537	1132808	29537	1132808	N	0	0	144	1	0
597B3	hind	6	8	p23.2	29582	2951410	29582	2951410	N	0	0	145	1	0
1014T2	xba	5	8	p23.2	87339	6154607	87340	6155384	N	1	0	146	1	0
1488T3	hind	5	8	p23.1	29954	12029748	29954	12029748	N	0	0	147	1	1

597B3	hind	6	8	p21.2	30324	24078370	30324	24078370	N	0	0	148	1	0
956T1	hind	6	8	p12	30515	32428808	30515	32428808	N	0	0	149	1	0
588T1	hind	6	8	p11.23	30659	38815153	30659	38815153	N	0	0	150	1	0
588T1	hind	6	8	q11.23	30853	53912575	30853	53912575	N	0	0	151	1	1
1466T2	hind	5	8	q13.1	31180	67079691	31181	67080307	N	1	1	152	4	0
334T2	hind	5	8	q13.1	31180	67079691	31181	67080307	N	1	1	152	4	0
1376T2	hind	5	8	q13.1	31180	67079691	31181	67080307	N	1	1	152	4	0
486T2	hind	5	8	q13.1	31180	67079691	31181	67080307	N	1	1	152	4	0
497B3	xba	5	8	q21.13	89141	84313002	89142	84313924	N	1	0	153	1	0
497B3	hind	4	8	q21.3	31780	90190503	31783	90271004	N	1	0	154	1	1
956T1	hind	6	8	q23.3	32389	116058634	32389	116058634	N	0	0	155	1	0
597B3	hind	6	8	q23.3	32430	116954963	32430	116954963	N	0	0	156	1	1
1013B1	hind	5	8	q24.11	32486	118662965	32487	118663176	N	1	1	157	3	0
1466T2	hind	5	8	q24.11	32486	118662965	32487	118663176	N	1	1	157	3	0
1488T3	hind	5	8	q24.11	32486	118662965	32487	118663176	N	1	1	157	3	0
486T2	hind	6	8	q24.21	32739	128341145	32739	128341145	N	0	0	158	1	0
597B3	xba	3	8	q24.23	90410	137748647	90423	137892295	N	1	1	159	5	1
869T1	xba	1	8	q24.23	90410	137748647	90423	137892295	N	1	1	159	5	1
869T1	hind	1	8	q24.23	32969	137758726	32974	137882148	N	1	1	159	5	1
1009B3	xba	1	8	q24.23	90412	137820222	90423	137892295	N	1	1	159	5	1
1229T2	xba	1	8	q24.23	90412	137820222	90423	137892295	N	1	1	159	5	1
1547T2	xba	1	8	q24.23	90412	137820222	90423	137892295	N	1	1	159	5	1
1052T1	xba	5	9	p24.2	90686	3786061	90686	3786061	N	0	0	160	1	0
1116B1	hind	3	9	p23	33427	10604555	33438	10765059	N	1	0	161	1	1
840B1	xba	4	9	p12	91775	42925816	91776	42930351	N	1	1	162	4	1
1052T1	xba	4	9	p12	91775	42925816	91776	42930351	N	1	1	162	4	1
1229T2	xba	4	9	p12	91775	42925816	91776	42930351	N	1	1	162	4	1
1376T2	xba	0	9	p12	91775	42925816	91776	42930351	N	1	1	162	4	1
334T2	hind	6	9	q21.11	34191	69089826	34191	69089826	N	0	0	163	1	0
588T1	hind	5	9	q21.2	34436	77099406	34437	77210528	N	1	0	164	1	0
1524T2	xba	5	9	q22.32	92227	93776663	92227	93776663	Y	0	0	165	1	0
597B3	hind	6	9	q31.1	34906	103639109	34906	103639109	N	0	0	166	1	0
1390	hind	0	9	q31.1	34952	105223959	34954	105240419	N	1	0	167	1	0
1430B1	xba	6	10	p13	93266	13871718	93266	13871718	N	0	0	168	1	0
956T1	xba	6	10	p12.2	93495	24542800	93495	24542800	N	0	0	169	1	0
1052T1	xba	5	10	p12.1	93525	25500769	93525	25500769	N	0	0	170	1	0
1466T2	hind	5	10	p11.1	36357	39114808	36357	39114808	N	0	1	171	2	1
486T2	hind	5	10	p11.1	36357	39114808	36357	39114808	N	0	1	171	2	1
1317T2	xba	5	10	q11.21	93857	43840664	93859	43952268	N	1	0	172	1	0
1401B3	xba	0	10	q21.1	94084	56014991	94085	56015341	Y	1	0	173	1	0
1116B1	hind	3	10	q22.1	36959	73708187	36966	74238077	N	1	0	174	1	1
1052T1	xba	6	10	q23.2	94839	86804174	94839	86804174	N	0	0	175	1	0
956T1	hind	5	10	q23.2	37236	88585539	37236	88585539	N	0	0	176	1	1

1542T1	hind	0	10	q23.31	37271	90515304	37273	90516359	N	1	0	177	1	0
956T1	hind	5	10	q23.31	37275	90589976	37276	90610172	N	1	0	178	1	0
588T1	hind	6	10	q25.1	37666	108675202	37666	108675202	Y	0	0	179	1	0
1104B2	hind	5	10	q26.13	38077	126075966	38077	126075966	N	0	0	180	1	0
588T1	hind	6	10	q26.13	38088	127376593	38088	127376593	N	0	0	181	1	0
840B1	xba	3	10	q26.3	95875	135164239	95879	135228726	N	1	1	182	5	1
1116B1	xba	3	10	q26.3	95875	135164239	95879	135228726	N	1	1	182	5	1
1153B3	xba	3	10	q26.3	95875	135164239	95879	135228726	N	1	1	182	5	1
1547T2	xba	3	10	q26.3	95875	135164239	95879	135228726	N	1	1	182	5	1
1376T2	hind	0	10	q26.3	38159	135211857	38159	135211857	N	0	1	182	5	1
1547T2	hind	5	10	q26.3	38159	135211857	38159	135211857	N	0	1	182	5	1
1390	hind	0	11	p15.4	38164	3359636	38164	3359636	N	0	0	183	1	1
1469B1	xba	3	11	p15.4	95901	4387439	95917	4499428	N	1	0	184	1	1
1390	xba	6	11	p15.4	95989	5860853	95989	5860853	N	0	0	185	1	1
1390	xba	6	11	p15.2	96129	12879375	96129	12879375	N	0	0	186	1	0
1395B3	hind	4	11	p13	38843	34612512	38845	34686463	N	1	0	187	1	0
597B3	hind	6	11	p12	38952	37835219	38952	37835219	N	0	0	188	1	0
1466T2	hind	5	11	p11.12	39135	51207947	39135	51207947	N	0	0	189	1	0
1014T2	xba	5	11	q12.1	97139	59115152	97139	59115152	N	0	0	190	1	0
862T1	xba	4	11	q14.1	97461	81554687	97463	81556741	N	1	0	191	1	0
1390	xba	6	11	q14.1	97523	83991296	97523	83991296	N	0	0	192	1	0
597B3	hind	6	11	q14.3	39670	88660169	39670	88660169	N	0	0	193	1	0
956T1	hind	5	11	q21	39817	94043963	39818	94053739	N	1	0	194	1	0
597B3	hind	6	11	q22.1	39934	98106915	39934	98106915	N	0	0	195	1	0
486T2	hind	6	11	q23.1	40274	110552629	40274	110552629	N	0	0	196	1	0
486T2	hind	6	11	q24.1	40496	122658485	40496	122658485	N	0	1	197	2	0
1390	hind	5	11	q24.1	40496	122658485	40496	122658485	N	0	1	197	2	0
588T1	hind	5	11	q24.2	40548	126107010	40548	126107010	N	0	0	198	1	0
597B3	hind	6	11	q25	40613	133028029	40613	133028029	N	0	0	199	1	0
956T1	hind	6	12	p12.3	40902	16182184	40902	16182184	Y	0	0	200	1	1
588T1	hind	6	12	p11.22	41261	29512321	41261	29512321	Y	0	0	201	1	0
956T1	hind	6	12	q13.11	41651	46575640	41651	46575640	N	0	0	202	1	0
1390	xba	5	12	q14.1	99918	58421669	99920	58460760	N	1	0	203	1	0
1062T1	xba	6	12	q14.1	99943	59831574	99943	59831574	N	0	0	204	1	1
810T1	xba	5	12	q21.2	100334	74579384	100334	74579384	N	0	0	205	1	0
588T1	hind	6	12	q21.33	42523	87828345	42523	87828345	N	0	0	206	1	0
1542T1	xba	0	12	q23.1	100909	98791609	100911	98791950	N	1	0	207	1	0
597B3	hind	5	12	q24.33	43176	128303727	43176	128303727	N	0	0	208	1	0
588T1	hind	6	13	q13.1	43530	31509971	43530	31509971	N	0	0	209	1	0
597B3	hind	6	13	q14.11	43897	42361754	43897	42361754	N	0	0	210	1	0
956T1	hind	6	13	q14.3	44118	51693873	44118	51693873	N	0	0	211	1	0
588T1	hind	6	13	q21.2	44310	59485089	44310	59485089	N	0	0	212	1	0
719T3	hind	0	13	q22.1	44693	72922299	44694	72922493	N	1	0	213	1	0

1052T1	xba	6	13	q31.2	103281	87190881	103281	87190881	N	0	0	214	1	0
588T1	hind	6	13	q32.1	45346	96361414	45346	96361414	N	0	0	215	1	0
956T1	hind	5	13	q33.3	45747	106977436	45750	107003052	N	1	0	216	1	0
597B3	xba	4	14	q11.2	104008	19285288	104012	19490525	N	1	1	217	10	1
1052T1	xba	3	14	q11.2	104008	19285288	104012	19490525	N	1	1	217	10	1
1399B3	xba	3	14	q11.2	104008	19285288	104012	19490525	N	1	1	217	10	1
1488T3	xba	1	14	q11.2	104008	19285288	104012	19490525	N	1	1	217	10	1
1547T2	xba	4	14	q11.2	104008	19285288	104012	19490525	N	1	1	217	10	1
486T2	xba	4	14	q11.2	104008	19285288	104012	19490525	N	1	1	217	10	1
697T2a	xba	3	14	q11.2	104008	19285288	104016	19560452	N	1	1	217	10	1
615B3	hind	4	14	q11.2	45852	19387587	45853	19440006	N	1	1	217	10	1
1116B1	hind	4	14	q11.2	45852	19387587	45853	19440006	N	1	1	217	10	1
810T1	hind	4	14	q11.2	45852	19387587	45853	19440006	N	1	1	217	10	1
1547T2	hind	5	14	q11.2	45852	19387587	45853	19440006	N	1	1	217	10	1
486T2	hind	4	14	q11.2	45852	19387587	45853	19440006	N	1	1	217	10	1
1153B3	xba	1	14	q11.2	104088	21703480	104103	21981289	N	1	1	217	2	1
1390	xba	6	14	q11.2	104103	21981289	104103	21981289	N	0	1	217	2	1
956T1	hind	6	14	q12	45974	25343544	45974	25343544	N	0	0	218	1	1
597B3	hind	6	14	q12	46057	27739847	46057	27739847	N	0	0	219	1	0
1052T1	xba	5	14	q21.2	104746	44308706	104746	44308706	N	0	0	220	1	1
1390	xba	6	14	q22.2	104967	54156618	104967	54156618	N	0	0	221	1	0
1430B1	xba	5	14	q23.1	105065	58179764	105066	58254253	N	1	0	222	1	0
956T1	hind	6	14	q23.3	47060	65715747	47060	65715747	N	0	0	223	1	0
1390	xba	5	14	q31.1	105575	80407821	105576	80433395	N	1	0	224	1	0
597B3	hind	6	14	q31.1	47362	81787150	47362	81787150	N	0	0	225	1	0
1401B3	xba	1	14	q31.3	105780	85967953	105792	86037625	N	1	0	226	1	0
1014T2	xba	5	14	q32.12	105901	91024899	105901	91024899	N	0	0	227	1	0
555B3	hind	3	14	q32.33	47775	104475429	47779	106241517	N	1	1	228	4	1
810T1	hind	3	14	q32.33	47776	105832953	47779	106241517	N	1	1	228	4	1
958B1	hind	4	14	q32.33	47778	106226043	47779	106241517	N	1	1	228	4	1
1399B3	hind	4	14	q32.33	47778	106226043	47779	106241517	N	1	1	228	4	1
497B3	xba	3	15	q11.2	106087	19208413	106091	19989036	N	1	1	228	26	1
615B3	xba	3	15	q11.2	106087	19208413	106091	19989036	N	1	1	228	26	1
538B3	xba	4	15	q11.2	106087	19208413	106091	19989036	N	1	1	228	26	1
840B1	xba	4	15	q11.2	106087	19208413	106091	19989036	N	1	1	228	26	1
956T1	xba	1	15	q11.2	106087	19208413	106091	19989036	N	1	1	228	26	1
1009B3	xba	3	15	q11.2	106087	19208413	106091	19989036	N	1	1	228	26	1
993B3	xba	1	15	q11.2	106087	19208413	106091	19989036	N	1	1	228	26	1
1013B1	xba	3	15	q11.2	106087	19208413	106091	19989036	N	1	1	228	26	1
1317T2	xba	1	15	q11.2	106087	19208413	106091	19989036	N	1	1	228	26	1
1399B3	xba	1	15	q11.2	106087	19208413	106091	19989036	N	1	1	228	26	1
1401B3	xba	3	15	q11.2	106087	19208413	106091	19989036	N	1	1	228	26	1
1488T3	xba	1	15	q11.2	106087	19208413	106091	19989036	N	1	1	228	26	1

869T1	xba	4	15	q11.2	106087	19208413	106091	19989036	N	1	1	228	26	1
334T2	xba	3	15	q11.2	106087	19208413	106099	20689912	N	1	1	228	26	1
1014T2	xba	3	15	q11.2	106087	19208413	106091	19989036	N	1	1	228	26	1
1229T2	xba	1	15	q11.2	106087	19208413	106091	19989036	N	1	1	228	26	1
597B3	hind	1	15	q11.2	47780	19852603	47784	19943075	N	1	1	228	26	1
615B3	hind	3	15	q11.2	47780	19852603	47784	19943075	N	1	1	228	26	1
538B3	hind	3	15	q11.2	47780	19852603	47784	19943075	N	1	1	228	26	1
840B1	hind	3	15	q11.2	47780	19852603	47784	19943075	N	1	1	228	26	1
956T1	hind	0	15	q11.2	47780	19852603	47784	19943075	N	1	1	228	26	1
1009B3	hind	4	15	q11.2	47780	19852603	47784	19943075	N	1	1	228	26	1
1542T1	hind	1	15	q11.2	47780	19852603	47784	19943075	N	1	1	228	26	1
993B3	hind	1	15	q11.2	47780	19852603	47784	19943075	N	1	1	228	26	1
1013B1	hind	3	15	q11.2	47780	19852603	47784	19943075	N	1	1	228	26	1
1317T2	hind	0	15	q11.2	47780	19852603	47784	19943075	N	1	1	228	26	1
1399B3	hind	1	15	q11.2	47780	19852603	47784	19943075	N	1	1	228	26	1
1433B3	hind	1	15	q11.2	47780	19852603	47784	19943075	N	1	1	228	26	1
810T1	hind	3	15	q11.2	47780	19852603	47784	19943075	N	1	1	228	26	1
869T1	hind	4	15	q11.2	47780	19852603	47784	19943075	N	1	1	228	26	1
1014T2	hind	4	15	q11.2	47780	19852603	47784	19943075	N	1	1	228	26	1
1229T2	hind	1	15	q11.2	47780	19852603	47784	19943075	N	1	1	228	26	1
719T3	hind	4	15	q11.2	47780	19852603	47784	19943075	N	1	1	228	26	1
1309T1	hind	1	15	q11.2	47780	19852603	47784	19943075	N	1	1	228	26	1
1538T3	hind	3	15	q11.2	47780	19852603	47784	19943075	N	1	1	228	26	1
1524T2	hind	3	15	q11.2	47780	19852603	47784	19943075	N	1	1	228	26	1
486T2	hind	3	15	q11.2	47780	19852603	47784	19943075	N	1	1	228	26	1
697T2a	hind	1	15	q11.2	47780	19852603	47784	19943075	N	1	1	228	26	1
1062T1	xba	5	15	q11.2	106107	21646501	106107	21646501	N	0	0	229	1	0
497B3	hind	3	15	q14	48144	37839046	48176	41999160	N	1	0	230	1	1
956T1	hind	5	15	q21.3	48501	53995275	48503	53998510	N	1	0	231	1	0
1390	xba	6	15	q21.3	106862	54133947	106862	54133947	N	0	0	232	1	0
597B3	hind	5	15	q22.2	48684	60359341	48684	60359341	N	0	0	233	1	0
1014T2	hind	6	15	q26.3	49210	98927879	49210	98927879	N	0	1	234	2	0
1309T1	hind	5	15	q26.3	49210	98927879	49210	98927879	N	0	1	234	2	0
956T1	xba	5	16	p13.3	107681	2747264	107682	2887015	N	1	0	235	1	0
588T1	hind	3	16	p11.2	49573	32411529	49578	35003380	N	1	1	236	2	1
1052T1	hind	4	16	p11.2	49573	32411529	49576	33463919	N	1	1	236	2	1
697T2a	xba	6	16	q22.1	108472	64398361	108472	64398361	N	0	0	237	1	0
1052T1	xba	6	16	q23.3	108814	81578029	108814	81578029	N	0	0	238	1	0
1116B1	xba	4	17	q12	109298	31462117	109301	31503652	N	1	1	239	2	1
1013B1	xba	4	17	q12	109298	31462117	109301	31503652	N	1	1	239	2	1
1469B1	hind	4	17	q21.31	50824	41550514	50827	41724181	N	1	1	239	2	1
1401B3	hind	5	17	q21.31	50826	41644356	50827	41724181	N	1	1	239	2	1
597B3	hind	6	17	q22	50953	49004409	50953	49004409	N	0	0	240	1	0

699T2a	hind	5	17	q23.3	51123	59640322	51123	59640322	N	0	0	241	1	0
810T1	xba	6	18	q21.1	110877	45572501	110877	45572501	N	0	0	242	1	0
956T1	hind	5	18	q22.2	52831	65887683	52832	65888907	N	1	0	243	1	0
1466T2	xba	5	18	q23	111664	71729056	111665	71729235	N	1	0	244	1	0
1052T1	xba	5	19	p13.2	111736	9464695	111737	9464751	N	1	0	245	1	0
1014T2	xba	5	19	p13.11	111765	16317260	111767	16452464	N	1	0	246	1	0
588T1	hind	6	19	q12	53210	35362367	53210	35362367	N	0	0	247	1	0
464T3	xba	6	19	q13.43	112069	61295070	112069	61295070	Y	0	0	248	1	0
1466T2	hind	3	19	q13.43	53394	63276577	53396	63458980	Y	1	1	249	4	0
1376T2	hind	4	19	q13.43	53394	63276577	53396	63458980	Y	1	1	249	4	0
515B3	hind	5	19	q13.43	53396	63458980	53396	63458980	Y	0	1	249	4	0
1153B3	hind	4	19	q13.43	53396	63458980	53396	63458980	Y	0	1	249	4	0
1401B3	xba	3	20	p13	112082	95685	112087	200721	N	1	0	250	1	1
699T2a	hind	5	20	p12.3	53482	5376300	53483	5376494	N	1	0	251	1	0
597B3	hind	6	20	p12.1	53686	13035735	53686	13035735	N	0	0	252	1	0
862T1	xba	4	20	p12.1	112479	15598897	112485	15600148	N	1	0	253	1	0
1166B3	xba	6	20	p11.23	112613	19982495	112613	19982495	N	0	0	254	1	0
597B3	hind	3	21	p11.2	54388	9928594	54390	10019412	N	1	0	255	1	1
597B3	hind	6	21	q21.2	54690	23311584	54690	23311584	Y	0	0	256	1	0
1401B3	hind	1	21	q21.2	54699	23661293	54707	23805469	Y	1	0	257	1	0
1390	xba	6	21	q21.2	113638	25067057	113638	25067057	Y	0	0	258	1	0
862T1	hind	0	21	q22.11	55038	34196798	55039	34197111	N	1	0	259	1	0
956T1	hind	6	21	q22.13	55110	37014570	55110	37014570	N	0	0	260	1	0
1399B3	xba	4	22	q11.1	114206	15263131	114207	15570421	N	1	1	261	2	1
1014T2	xba	5	22	q11.1	114207	15570421	114207	15570421	N	0	1	261	2	1
597B3	hind	6	22	q11.22	55315	20776031	55315	20776031	N	0	0	262	1	0
946B3	hind	4	23	p22.33	55700	677050	55700	677050	N	0	1	263	3	1
956T1	hind	3	23	p22.33	55700	677050	55716	3316027	N	1	1	263	3	1
334T2	hind	0	23	p22.33	55700	677050	55700	677050	N	0	1	263	3	1
956T1	hind	6	23	p22.33	55717	3686178	55717	3686178	N	0	0	264	1	0
588T1	hind	5	23	p22.31	55759	7649712	55760	7650390	N	1	0	265	1	0
464T3	xba	5	23	p22.11	114736	22594910	114737	22597318	N	1	0	266	1	0
1052T1	xba	3	23	p21.1	114897	35002772	114933	40393087	Y	1	1	267	2	1
1081B2	xba	4	23	p11.4	114919	37926792	114921	37978787	Y	1	1	268	2	0
1052T1	xba	4	23	p11.4	114934	40958413	114936	41146967	Y	1	0	269	1	0
956T1	xba	5	23	p11.3	114964	45448103	114964	45448103	Y	0	0	270	1	0
1065B3	xba	5	23	p11.23	114979	48918038	114979	48918038	Y	0	0	271	1	1
497B3	xba	3	23	q12	115010	64965125	115018	65600282	N	1	0	272	1	0
956T1	hind	5	23	q21.1	56225	77347614	56226	77426206	N	1	1	273	4	0
848B1	hind	4	23	q21.1	56225	77347614	56226	77426206	N	1	1	273	4	0
1466T2	hind	5	23	q21.1	56225	77347614	56226	77426206	N	1	1	273	4	0
697T2a	hind	5	23	q21.1	56225	77347614	56226	77426206	N	1	1	273	4	0
1390T3	xba	6	23	q23	115324	114623485	115324	114623485	N	0	0	274	1	0

1081B2	xba	3	23	q24	115368	118876414	115378	120796067	N	1	0	275	1	1
956T1	hind	6	23	q25	56587	122556844	56587	122556844	N	0	0	276	1	0
956T1	hind	3	23	q25	56588	122765447	56601	124002324	N	1	0	277	1	0
597B3	hind	6	23	q27.1	56721	139718845	56721	139718845	N	0	0	278	1	1
956T1	hind	6	23	q28	56819	147227404	56819	147227404	N	0	0	279	1	0
1542T1	xba	3	23	q28	115720	149973397	115737	154409808	N	1	0	280	1	1

[a] CNV regions indentified per sample

[b] CN, copy number of the region per samples

[c] StartSNP, SNP index where the CN region begins; StartPos, chromosome position where the CNV region begins

[d] Y, CNV regions located in a known AD linked region

[e] 1, CNV regions containing >1 SNP

[f] 1, CNV regions indentified by >1 sample

[g] index for CNV regions

[h] Number of samples for each region

[i] 1, region is known in the published database

Table 11. Number of Unique CNV regions located with in known AD linkage regions

	# of CNV regions				
linked regions[a]	**total #**	**> 1 SNP[b]**	**> 1 sample[c]**	**coding[d]**	**Published region[e]**
1p36	2	1	0	Y	0
1q23-31	0	0	0	NA	0
2p23-24	1	1	0	N	0
4q35	1	0	0	N	0
5p13-15	1	0	0	Y	1
6p21	2	1	2	Y	1
6q15-16	2	1	0	N	0
6q25-27	3	1	0	Y	0
9p21	0	0	0	NA	0
9q22	1	0	0	N	0
10q21-22	1	1	0	Y	0
10q25	1	0	0	N	0
12p11-12	2	0	0	N	1
19q13	3	2	4	Y	0
21q21-22	3	1	0	N	0
Xp11-21	4	3	2	N	2

[a] Known AD linkage regions [69]
[b] 1, CNV regions containing >1 SNP
[c] 1, CNV regions indentified by >1 sample
[d] Y, CNV region is located in a coding region
[e] 1, region is known in the published database

Table 12. Comparison of CNV regions with known CNV regions[a].

Chr	Start SNP[b]	End SNP	Variation ID[c]	Startd	Stop	KnownGenes
1q44	4521	4522	3338	242733187	243066985	TFB2M\tC1orf71\tSMYD3
chr2_cent	64104	64104	528	95520061	95773837	TRIM43
2q37.3	9604	9609	3410	242139945	242221648	STK25\tBOK\tFARP2
3p26.3	9610	9623	3412	46156	773503	CHL1
3p26.3	9689	9690	3414	1400003	1563295	CNTN6
3q12.2	69181	69181	3440	102096077	102259187	ABI3BP
3q23	69972	69972	2476	143239239	143577329	MGC40579\tXRN1
3q23	69972	69972	3451	143211568	143612534	MGC40579\tTFDP2\tXRN1
3q29	13560	13561	568	196922245	198866401	FLJ25996\tMFI2\tMUC20\tLOC440993\tPCYT1A\tPAK2\tDLG1\tBDH1\tPIGX\tMUC4\tNCBP2\tSENP5\tRNF168\tOSTalpha\tTFRC\tZDHHC19\tLRRC33\tC3orf34\tTNK2\tMGC33212\tTM4SF19\tPIGZ\tWDR53
3q29	13560	13561	43	198380968	198552148	DLG1
3q29	13560	13561	3473	198343363	198986178	DLG1\tBDH1\tFYTTD1
5q31.3	21194	21194	714	140516163	140568127	PCDHB11\tPCDHB16\tPCDHB7\tPCDHB8\tPCDHB9\tPCDHB10
5q31.3	21194	21194	3578	140137124	140808866	PCDHAC1\tSLC25A2\tPCDHB16\tPCDHB2\tPCDHAC2\tPCDHB7\tPCDHB15\tPCDHGA10\tPCDHGB2\tPCDHGA12\tPCDHB9\tPCDHGA9\tPCDHGA8\tPCDHGB3\tPCDHGB6\tPCDHGA5\tPCDHA6\tPCDHA8\tPCDHA13\tPCDHB5\tPCDHB13\tPCDHA3\tPCDHGA4\tPCDHA4\tPCDHGA7\tPCDHB4\tPCDHB11\tPCDHA2\tPCDHGB5\tPC
6p21.1	23026	23040	3609	41696767	41807431	MDFI\tTFEB
6p21.1	23026	23040	2626	41706502	41749467	MDFI
6q22.31	82354	82357	3636	122765634	122925836	HSF2\tSERINC1\tPKIB
6q24.2	25514	25516	2656	143861702	143866797	FUCA2
7q11.21	27674	27685	3683	65845580	66303403	SBDS\tFLJ10099\tRSAFD1
8p23.1	29954	29954	2736	11882733	12646342	FAM86B1\tDEFB134\tDUB3\tLONRF1
8p23.1	29954	29954	3720	11778967	12611672	DEFB134\tDEFB136\tDUB3\tFAM86B1\tDEFB137
8p23.1	29954	29954	349	11908066	12332070	FAM86B1\tDUB3
8q11.23	30853	30853	3733	53823346	54105603	GPR7
9p12	91775	91776	2818	42050602	43951965	ZNF658B
10q22.1	36959	36966	3809	74105644	74277126	C10orf42
10q23.2-10q23.31	37236	37236	3815	88505038	89299742	C10orf116\tBMPR1A\tMMRN2\tSNCG\tMINPP1\tGLUD1\tFAM35A
10q26.3	95875	95879	3830	135111207	135279389	SPRN\tSYCE1\tFLJ44653\tMTG1\tCYP2E1

10q26.3	95875	95879	2896	134907049	135321566	DRD1IP\tSPRN\tVENTX\tSYCE1\tKNDC1\tMTG1\tCYP2E1\tPAOX\tC10orf125\tFLJ44653\tADAM8\tZNF511\tECHS1\tUTF1\tPRAP1\tTUBGCP2
10q26.3	95875	95879	370	135130574	135413628	SYCE1\tCYP2E1\tFLJ44653\tDUX4
10q26.3	95875	95879	2162	135117752	135279389	SPRN\tSYCE1\tMTG1\tCYP2E1\tFLJ44653
11p15.4	38164	38164	3833	3193088	3758005	NUP98\tZNF195\tART1\tART5\tC11orf36\tCHRNA10
11p15.4	95901	95917	2899	4466861	4519034	OR52K1
11p15.4	95989	95989	2903	5824631	5964189	OR52E5\tOR52L1\tOR56A3\tOR52E4\tOR52E8
11p15.4	95989	95989	3836	5686953	5927293	OR52N4\tOR52N2\tOR56A3\tOR56B1\tOR52E4\tOR52E5\tOR52E6\tOR52N5\tTRIM22\tOR52N1\tOR52E8
14q11.2	104008	104012	2175	19440982	19574860	OR4K5\tOR4K13\tOR4K15\tOR4K1\tOR4K14
chr14_cent-14q11.2	104008	104012	3929	18732531	19626027	OR4K5\tOR4K13\tOR4K15\tACTBL1\tOR4Q3\tOR4L1\tOR4K14\tLOC440157\tOR4K2\tOR4M1\tOR4K1\tOR4N2
14q11.2	104008	104012	376	19273689	19767232	OR4K5\tOR4K13\tOR4K15\tOR11H6\tOR4K2\tOR4M1\tOR4Q3\tOR4L1\tOR4K1\tOR4K14\tOR4N2\tOR4N5\tOR11G2\tOR4K17
14q11.2	104008	104012	3036	19272965	19608948	OR4K5\tOR4K13\tOR4K15\tOR4K2\tOR4M1\tOR4Q3\tOR4L1\tOR4K1\tOR4K14\tOR4N2
14q11.2	45852	45853	316	19267456	19468888	OR4K5\tOR4K2\tOR4M1\tOR4Q3\tOR4N2
14q32.33	47775	47779	2176	105294030	105477141	KIAA0125
14q32.33	47775	47779	3950	104584391	105020471	PACS2\tJAG2\tBTBD6\tMTA1\tGPR132\tCRIP2\tNUDT14\tBRF1
15q11.2	106087	106099	1281	20484859	20567999	NIPA2\tCYFIP1
15q11.2	106087	106091	443	18870124	20077222	OR4M2\tLOC283755\tLOC400968\tPOTE15\tOR4N4
15q11.2	106087	106091	2182	19808433	19975452	OR4M2\tOR4N4
15q11.2	106087	106099	352	20459937	21183722	NIPA1\tGOLGA8E\tNIPA2\tCYFIP1
15q11.2	106087	106091	3070	18403665	21241985	NIPA1\tGOLGA8E\tNIPA2\tTUBGCP5\tCYFIP1\tOR4M2\tLOC283755\tFLJ36144\tLOC400968\tPOTE15\tOR4N4
15q15.1	48144	48176	3959	38619837	38782403	CCDC32\tCASC5\tRAD51\tRPUSD2
15q15.3	48144	48176	3960	41610450	41935933	STRC\tMAP1A\tSERINC4\tHISPPD2A\tMFAP1\tCKMT1A\tCATSPER2\tCKMT1B\tSERF2\tELL3\tHYPK\tPDIA3\tWDR76
15q15.3	48144	48176	1282	41676268	41726934	STRC\tCATSPER2\tCKMT1B
15q15.1	48144	48176	785	40121667	40290274	PLA2G4D\tVPS39\tPLA2G4F\tTMEM87A
15q15.3	48144	48176	3079	41619215	41845926	STRC\tCATSPER2\tCKMT1B\tPDIA3\tHISPPD2A\tCKMT1A
15q15.1	48144	48176	3078	39834465	39909909	MAPKBP1\tPLA2G4B
16p11.2	49573	49576	2214	32541841	32746227	TP53TG3
16p11.2	49573	49576	2212	32646445	32828220	FLJ43855
16p11.2	49573	49576	324	32082890	33533065	TP53TG3\tFLJ43855
16p11.2	49573	49576	4002	31658070	34219714	LOC124411\tTP53TG3\tZNF267\tFLJ43855

17q12	109298	109301	3142	31429427	32016863	TBC1D3C\tCCL4L1\tZNHIT3\tCCL4L2\tCCL3L1\tCCL3L3\tPIGW\tCCL4\tZNF403\tCCL3\tMYOHD1\tTBC1D3B
17q12	109298	109301	4031	31309398	31981395	CCL16\tTBC1D3C\tCCL3\tMYOHD1\tCCL18\tTBC1D3B\tCCL15\tCCL14\tCCL4L1\tZNHIT3\tCCL4L2\tCCL3L1\tCCL3L3\tCCL23\tPIGW\tCCL4\tZNF403
17q21.31	50824	50827	327	41515374	42125543	LOC474170\tNSF\tARL17P1\tKIAA1267\tLRRC37A
17q21.31	50824	50827	2227	41572525	41734030	KIAA1267\tLRRC37A
17q21.31	50824	50827	3147	41518102	42223353	LOC474170\tNSF\tARL17P1\tKIAA1267\tLRRC37A\tWNT3
17q21.32 - 17q21.31	50824	50827	4038	41439751	42632332	LOC474170\tWNT9B\tGOSR2\tARL17P1\tLRRC37A\tCDC27\tRPRML\tNSF\tKIAA1267\tMAPT\tWNT3
20p13	112082	112087	2249	37211	226388	ZCCHC3\tC20orf96\tRP5-1103G7.6\tDEFB127\tDEFB129\tDEFB126
21p11.2	54388	54390	1305	9978594	9986010	TPTE
chr21_cent-21p11.2	54388	54390	3224	9941889	10105718	BAGE\tBAGE2\tTPTE\tBAGE5\tBAGE4\tBAGE3
chr22_cent	114206	114207	4115	14509865	15652016	XKR3\tCESK1\tOR11H1
Xp22.33	55700	55716	3260	2696752	3169853	ARSH\tXG\tARSF\tGYG2\tARSE\tARSD
Xp22.33	55700	55716	4142	1865728	3187726	XG\tGYG2\tARSE\tCD99\tARSH\tARSF\tZBED1\tARSD
Xp11.23	114979	114979	2278	48817974	49065087	GAGE4\tGAGE7B\tCCDC22\tGAGE5\tCACNA1F\tPPP1R3F\tGAGE8\tGAGE1\tGAGE6\tGAGE2\tGAGE7\tFOXP3
Xq24	115368	115378	668	119789968	119818984	RP6-166C19.1
Xq24	115368	115378	4168	119183307	119364831	LAMP2\tATP1B4\tFAM70A
Xq27.2-Xq27.1	56721	56721	4175	139711783	140559532	SPANXD\tSPANXB1\tSPANXA1\tLDOC1\tSPANXA2\tSPANXB2\tSPANXE\tSPANXC
Xq28	115720	115737	3259	151900926	152045136	PNMA6A\tMAGEA1
Xq28	115720	115737	3273	151572622	151688320	MAGEA3\tCETN2\tCSAG1\tCSAG2\tNSDHL\tMAGEA2B\tMAGEA12\tCSAG3A\tMAGEA2
Xq28	115720	115737	673	153081928	153109733	FLNA
Xq28	115720	115737	672	151934123	152001387	MAGEA1
Xq28	115720	115737	1870	154338367	154341268	TMLHE
Xq28	115720	115737	674	153126494	153149694	RPL10\tEMD
Xq28	115720	115737	4178	151881624	151947652	PNMA6A\tPNMA3

[a] Known CNV regions were obtained from Database for Genomic Variants ([137], http://projects.tcag.ca/variation/) and compared with regions found in this study. This gene list has been filtered to include only unique genes.

[b] StartSNP, SNP index where the CN region begins

[c] Index ID of CNV region in the Database for Genomic Variants

Table 13. Comparison of expression probe sets and CNV regions[a]

sample	Chr	Probe Set ID	Gene.Title	diff.expr[b]	coding[c]	published[d]
588T1	chr1p36.22	226532_at	Leucine zipper and CTNNBIP1 domain containing	0	1	0
597B3	chr1p36.12	201935_s_at	eukaryotic translation initiation factor 4 gamma, 3	0	1	0
588T1	chr1p33	218080_x_at	Fas (TNFRSF6) associated factor 1	0	1	0
597B3	chr1p32-p31	205805_s_at	receptor tyrosine kinase-like orphan receptor 1	0	1	0
597B3	chr1p32-p31	211057_at	receptor tyrosine kinase-like orphan receptor 1 /// receptor tyrosine kinase-like orphan receptor 1	0	1	0
1547T2	chr1p31.1	1553194_at	neuronal growth regulator 1	1	1	0
1488T3	chr1q25.2	1553376_a_at	chromosome 1 open reading frame 125	0	0	0
1466T2	chr1q44	233578_at	chromosome 1 open reading frame 101	0	1	1
810T1	chr2p16.3	232364_at	F-box protein 11	0	1	0
1430B1	chr2p13.3	1570337_at	folliculogenesis specific basic helix-loop-helix	0	1	0
1430B1	chr2p13.3	1552410_at	C-type lectin domain family 4, member F	0	1	0
1430B1	chr2p13	220428_at	CD207 molecule, langerin	0	1	0
1430B1	chr2q33.1	202918_s_at	preimplantation protein 3	1	1	0
1430B1	chr2q11.2	1560871_a_at	Similar to regulatory factor X domain containing 1	0	1	0
597B3	chr2q21	1557386_at	Lactase	0	0	0
1052T1	chr2q31.1	239798_at	Pyruvate dehydrogenase kinase, isozyme 1	0	0	0
486T2	chr2q31.2	224002_s_at	FK506 binding protein 7	0	1	0
597B3	chr2q32.2	213374_x_at	3-hydroxyisobutyryl-Coenzyme A hydrolase	0	0	0
956T1	chr2q36.3	226281_at	delta-notch-like EGF repeat-containing transmembrane	1	1	0
1433B3	chr2q37	235624_at	High density lipoprotein binding protein (vigilin)	0	1	1
1166B3	chr2q37.3	225652_at	FERM, RhoGEF and pleckstrin domain protein 2	0	1	1
1166B3	chr2q37	200778_s_at	septin 2	0	1	1
1166B3	chr2q37.3	201314_at	serine/threonine kinase 25 (STE20 homolog, yeast)	1	1	1
1166B3	chr2q37.3	1553046_s_at	galactose-3-O-sulfotransferase 2	0	1	1
1433B3	chr2q37	200015_s_at	septin 2 /// septin 2	0	1	1
1433B3	chr2q37.3	1559038_at	similar to septin 2	0	1	1
1433B3	chr2q37.3	1559617_at	Sialidase 4	0	1	1
1166B3	chr2q37.3	241918_at	transmembrane protein 16G	0	1	1
1433B3	chr2q37.3	219674_s_at	hypothetical protein PRO2900	0	1	1
1166B3	chr2q37.3	223349_s_at	BCL2-related ovarian killer	0	1	1
615B3	chr3q13.33	228940_at	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 4, 15kDa /// similar to NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 4, 15kDa	0	1	1
334T2	chr3q29	1556103_at	Tyrosine kinase, non-receptor, 2	0	1	0

956T1	chr3p21.31	219306_at	kinesin family member 15	0	1	0
1014T2	chr3p14.3	209285_s_at	chromosome 3 open reading frame 63	0	1	0
597B3	chr3p14.2	206492_at	fragile histidine triad gene	1	1	0
1052T1	chr3q13.3	209945_s_at	glycogen synthase kinase 3 beta	0	1	0
588T1	chr3q29	213184_at	SUMO1/sentrin specific peptidase 5	0	1	1
588T1	chr3q28-q29	243629_x_at	Antigen p97 (melanoma associated) identified by monoclonal antibodies 133.2 and 96.5	0	1	1
588T1	chr3q29	229703_at	Discs, large homolog 1 (Drosophila)	0	1	1
588T1	chr3q29	225657_at	hypothetical protein BC007882	1	1	1
588T1	chr3q29	201521_s_at	nuclear cap binding protein subunit 2, 20kDa	0	1	1
588T1	chr3q29	220041_at	phosphatidylinositol glycan, class Z	0	1	1
588T1	chr3q29	208875_s_at	p21 (CDKN1A)-activated kinase 2	0	1	1
1052T1	chr4q13.1	209866_s_at	latrophilin 3	0	1	0
1014T2	chr4q22	221364_at	glutamate receptor, ionotropic, delta 2	0	1	0
697T2a	chr4q22.3	241479_at	Hypothetical LOC441031	0	1	0
956T1	chr4q25-q26	220792_at	PR domain containing 5	0	1	0
1014T2	chr4q26	201701_s_at	progesterone receptor membrane component 2	1	1	0
1062T1	chr4q31.21	223878_at	inositol polyphosphate-4-phosphatase, type II, 105kDa	0	1	0
1430B1	chr5q11.2	208394_x_at	endothelial cell-specific molecule 1	0	1	0
946B3	chr5q13.2	1557278_s_at	Transportin 1	0	1	0
597B3	chr5q14.1	224167_at	spermatogenic leucine zipper 1	0	1	0
597B3	chr5p15.2-q14.3	1554638_at	zinc finger, FYVE domain containing 16	0	1	0
1309T1	chr5q22.2	220120_s_at	erythrocyte membrane protein band 4.1 like 4A	0	1	0
956T1	chr5q23.3	242100_at	chondroitin sulfate synthase 3	0	0	0
588T1	chr5q31	221450_x_at	protocadherin beta 13	0	0	1
588T1	chr5q31	232099_at	protocadherin beta 16	0	0	1
588T1	chr5q31	231725_at	protocadherin beta 2	1	1	1
588T1	chr5q31	231738_at	protocadherin beta 7	0	0	1
1014T2	chr5q31-q33	216939_s_at	5-hydroxytryptamine (serotonin) receptor 4	0	1	0
956T1	chr6p25.2	1558882_at	similar to HIV TAT specific factor 1; cofactor required for Tat activation of HIV-1 transcription	0	1	0
597B3	chr6p22.3-p22.2	206017_at	KIAA0319	0	1	0
697T2a	chr6p21.3-p21.2	202530_at	mitogen-activated protein kinase 14	0	1	0
497B3	chr6p21.31	223516_s_at	chromosome 6 open reading frame 49	1	1	1
497B3	chr6p21.1	206961_s_at	Trf (TATA binding protein-related factor)-proximal homolog (Drosophila)	0	1	1
497B3	chr6p21.1	1561959_x_at	KIAA0240	0	1	1
497B3	chr6p21	1562028_at	Cyclin D3	0	1	1

497B3	chr6p21.1	238028_at	similar to AI661453 protein	0	1	1
497B3	chr6p21.1-p12.1	224247_s_at	mitochondrial ribosomal protein S10	1	1	1
497B3	chr6p21.1	217493_x_at	natural cytotoxicity triggering receptor 2	0	1	1
497B3	chr6p21	1564029_at	ubiquitin specific peptidase 49	0	1	1
497B3	chr6p21.1	235528_at	guanylate cyclase activator 1B (retina)	0	1	1
497B3	chr6p21	221866_at	transcription factor EB	0	1	1
497B3	chr6pter-p12.1	202495_at	tubulin-specific chaperone c	1	1	1
497B3	chr6p21.1	203612_at	bystin-like	0	1	1
497B3	chr6p21.1	206062_at	guanylate cyclase activator 1A (retina)	0	1	1
497B3	chr6p21.2-p12.3	206625_at	retinal degeneration, slow	0	1	1
497B3	chr6p21.3-p21.1	205261_at	progastricsin (pepsinogen C)	0	1	1
497B3	chr6p21	205375_at	MyoD family inhibitor	0	1	1
497B3	chr6p21.1	1552494_at	taube nuss homolog (mouse)	0	1	1
497B3	chr6p21.1	219907_at	fibroblast growth factor receptor substrate 3	0	1	1
497B3	chr6p21.1	1561770_at	Ribosomal protein L7-like 1	0	1	1
497B3	chr6p21.1-p12.1	239783_at	Transcriptional regulating factor 1	0	1	1
497B3	chr6p21.1	229763_at	forkhead box P4	0	1	1
497B3	chr6p21.1	215558_at	Ubiquitin protein ligase E3 component n-recognin 2	0	1	1
497B3	chr6p21.1	229914_at	FLJ38717 protein	1	1	1
597B3	chr6q12-q13	1552922_at	regulating synaptic membrane exocytosis 1	0	1	0
1430B1	chr6q22.31	208671_at	serine incorporator 1	1	1	1
1430B1	chr6q22.31	209657_s_at	heat shock transcription factor 2	0	1	1
588T1	chr6q22-q23	203786_s_at	tumor protein D52-like 1	1	1	0
956T1	chr6q23.1-23.3	243697_at	Monooxygenase, DBH-like 1	0	1	0
956T1	chr6q23-q24	1560346_at	HBS1-like (S. cerevisiae)	0	0	0
956T1	chr6q23-q24	218603_at	headcase homolog (Drosophila)	0	1	0
956T1	chr6q24	223120_at	fucosidase, alpha-L- 2, plasma	0	1	1
956T1	chr6q24.2	1556859_a_at	hypothetical protein LOC285740	0	1	1
956T1	chr6q25.1	220329_s_at	chromosome 6 open reading frame 96	1	1	0
956T1	chr7p	207775_at	hypothetical protein MGC4859 similar to HSPA8	1	1	0
862T1	chr7p15.3	206663_at	Sp4 transcription factor	0	1	0
1062T1	chr7p15.2-p15.1	225798_at	juxtaposed with another zinc finger gene 1	1	1	0
956T1	chr7p13	1569659_at	HECT, C2 and WW domain containing E3 ubiquitin protein ligase 1	0	1	0
956T1	chr7p12-p14	202693_s_at	serine/threonine kinase 17a (apoptosis-inducing)	0	1	0
956T1	chr7p13	209445_x_at	hypothetical protein FLJ10803	1	1	0
497B3	chr7p22	210933_s_at	fascin homolog 1, actin-bundling protein (Strongylocentrotus purpuratus)	0	1	1
497B3	chr7q11.21	218008_at	hypothetical protein FLJ10099	0	1	1

497B3	chr7q11.21	218310_at	RAB guanine nucleotide exchange factor (GEF) 1	0	1	1
497B3	chr7q11.21	1561964_at	hypothetical LOC441245	0	1	1
497B3	chr7q11-q22	215192_at	postmeiotic segregation increased 2-like 4	0	1	1
497B3	chr7q11.21	213474_at	Potassium channel tetramerisation domain containing 7	0	1	1
497B3	chr7q11.21 /// chr7q11.23	1554089_s_at	Shwachman-Bodian-Diamond syndrome /// Shwachman-Bodian-Diamond syndrome pseudogene	0	1	1
497B3	chr7q11.21	222669_s_at	Shwachman-Bodian-Diamond syndrome	1	1	1
497B3	chr7q11.21	239896_at	RAB guanine nucleotide exchange factor (GEF) 1 pseudogene	0	1	1
497B3	chr7q11.21	238921_at	hypothetical protein LOC641767 /// hypothetical protein LOC644794 /// hypothetical protein LOC649972	0	1	1
497B3	chr7p11.2-q11.2	231521_at	Stromal antigen 3-like	0	1	1
597B3	chr7q11	223885_at	calneuron 1	0	1	0
993B3	chr7q21.13	219455_at	hypothetical protein FLJ21062	0	1	0
515B3	chr7q31.32	1568924_a_at	hypothetical protein FLJ35834	0	1	0
515B3	chr7q32	215850_s_at	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 5, 13kDa	0	1	0
1390	chr7q31.3-q32.1	216992_s_at	glutamate receptor, metabotropic 8	0	1	0
956T1	chr7q32.1	212814_at	KIAA0828 protein	0	1	0
1390	chr7q33-q34	211737_x_at	pleiotrophin (heparin binding growth factor 8, neurite growth-promoting factor 1) /// pleiotrophin (heparin binding growth factor 8, neurite growth-promoting factor 1)	0	1	0
1390	chr7q33-q34	209465_x_at	pleiotrophin (heparin binding growth factor 8, neurite growth-promoting factor 1)	0	1	0
1116B1	chr7q35-q36	219301_s_at	contactin associated protein-like 2	1	1	0
597B3	chr8p23.2	1553405_a_at	CUB and Sushi multiple domains 1	0	1	0
956T1	chr8p21-p12	208232_x_at	neuregulin 1	1	1	0
588T1	chr8p11	1554690_a_at	transforming, acidic coiled-coil containing protein 1	0	1	0
1430B1	chr10p13	208476_s_at	FERM domain containing 4A	1	1	0
956T1	chr10p12.1	1560115_a_at	KIAA1217	0	1	0
1401B3	chr10p11.2-q21	1560330_at	protocadherin 15	0	1	0
1116B1	chr10q22.1	225320_at	coiled-coil domain containing 109A	0	1	1
1116B1	chr10q22.1	214338_at	DnaJ (Hsp40) homolog, subfamily B, member 12	0	1	1
1116B1	chr10q22.1	216903_s_at	calcium binding atopy-related autoantigen 1	0	1	1
1116B1	chr10q22.1	243335_at	Chromosome 10 open reading frame 42	0	1	1
956T1	chr10q22.3	243275_at	Bone morphogenetic protein receptor, type IA	1	0	1
588T1	chr10q26.2	1557591_at	hypothetical protein LOC283038	0	1	0
1547T2	chr10q26.3	233084_s_at	synaptonemal complex central element protein 1	0	1	1
840B1	chr10q24.3-qter	1431_at	cytochrome P450, family 2, subfamily E, polypeptide 1	1	1	1
1390	chr11p15.5	204234_s_at	zinc finger protein 195	0	0	1

1469B1	chr11p15.4	232829_at	olfactory receptor, family 52, subfamily K, member 3 pseudogene	0	1	1
1390	chr11p15.2	1553322_s_at	TEA domain family member 1 (SV40 transcriptional enhancer factor)	0	1	0
1395B3	chr11p12	232360_at	ets homologous factor	0	1	0
1014T2	chr11q12-q13	201800_s_at	oxysterol binding protein	0	1	0
1390	chr11q14.1	206253_at	discs, large homolog 2, chapsyn-110 (Drosophila)	0	1	0
597B3	chr11q14-q21	206630_at	tyrosinase (oculocutaneous albinism IA)	0	1	0
956T1	chr12q13.11	204253_s_at	vitamin D (1,25- dihydroxyvitamin D3) receptor	0	1	0
1390	chr12q13	210807_s_at	solute carrier family 16 (monocarboxylic acid transporters), member 7	0	1	0
588T1	chr13q13.1	204072_s_at	furry homolog (Drosophila)	0	1	0
597B3	chr13q13.3	227609_at	epithelial stromal interaction 1 (breast)	0	1	0
956T1	chr13q14.3	1556876_s_at	TPTE pseudogene 1	0	1	0
588T1	chr13q21.2	220997_s_at	diaphanous homolog 3 (Drosophila) /// diaphanous homolog 3 (Drosophila)	0	1	0
588T1	chr13q21.2	242102_at	Diaphanous homolog 3 (Drosophila)	0	1	0
1153B3	chr14q11.2	211902_x_at	T cell receptor alpha locus	0	1	1
1153B3	chr14q11.2 /// chr14q11	209671_x_at	T cell receptor alpha locus /// T cell receptor alpha locus /// T cell receptor alpha constant /// T cell receptor alpha constant	0	1	1
1153B3	chr14q11.2 /// chr14q11	210972_x_at	T cell receptor alpha locus /// T cell receptor delta variable 2 /// T cell receptor alpha variable 20 /// T cell receptor alpha constant	0	1	1
1153B3	chr14q11	209670_at	T cell receptor alpha constant /// T cell receptor alpha constant	0	1	1
1153B3	chr14q11.2	234388_at	T cell receptor alpha locus /// T cell receptor alpha chain /// Rearranged T-cell receptor alpha-chain mRNA, variable region	0	1	1
1153B3	chr14q11.2	217143_s_at	T cell receptor alpha locus /// T cell receptor delta locus	0	1	1
1153B3	chr14q11.2	217056_at	T cell receptor alpha locus /// T cell receptor, clone IGRA15 /// T-cell receptor alpha chain V-region /// TCRA PS7 mRNA	0	1	1
1390	chr14q22.2	230503_at	Sterile alpha motif domain containing 4A	0	0	0
1430B1	chr14q23.1	219179_at	dapper, antagonist of beta-catenin, homolog 1 (Xenopus laevis)	0	1	0
1390	chr14q31.1	233859_at	chromosome 14 open reading frame 145	0	1	0
1014T2	chr14q31-q32	210760_x_at	Thyroid hormone receptor interactor 11	0	1	0
1014T2	chr14q32.12	220369_at	KIAA2010	0	1	0
555B3	chr14q32.33	211636_at	immunoglobulin heavy constant alpha 1 /// immunoglobulin heavy constant alpha 1 /// immunoglobulin heavy constant alpha 2 (A2m marker) /// immunoglobulin heavy constant alpha 2 (A2m marker) /// immunoglobulin heavy constant gamma 1 (G1m marker) /// immuno	0	1	1
555B3	chr14q32.33	228558_at	chromosome 14 open reading frame 80	0	1	1
555B3	chr14q32.33	211647_x_at	Immunoglobulin heavy constant gamma 1 (G1m marker) /// Immunoglobulin heavy constant gamma 1 (G1m marker)	0	1	1
555B3	chr14q32.32-q32.33	211640_x_at	Immunoglobulin heavy variable 1-69 /// Immunoglobulin heavy variable 1-69	0	1	1

555B3	chr14q32.33	1558437_at	Immunoglobulin epsilon chain constant region=membrane-bound form {M:A, alternatively spliced, exon C4, membrane domain exon 1} [human, B cell myeloma U-266, mRNA Partial, 230 nt] /// Epsilon , IgE=membrane-bound IgE, epsilon m/s isoform {alternative splic	0	1	1
555B3	chr14q32.33	211637_x_at	Hypothetical protein LOC90925 /// Hypothetical protein LOC90925	0	1	1
555B3	chr14q32.3	211783_s_at	metastasis associated 1 /// metastasis associated 1	0	1	1
555B3	chr14q32.33	212827_at	immunoglobulin heavy constant mu /// immunoglobulin heavy constant mu	0	1	1
555B3	chr14q32.33	212778_at	phosphofurin acidic cluster sorting protein 2	0	1	1
555B3	chr14q32.3	1568929_at	Metastasis associated 1	0	1	1
555B3	chr14q32.33	211835_at	immunoglobulin heavy locus /// immunoglobulin heavy constant alpha 1 /// immunoglobulin heavy constant alpha 2 (A2m marker) /// immunoglobulin heavy constant delta /// immunoglobulin heavy constant gamma 1 (G1m marker) /// immunoglobulin heavy constant ga	0	1	1
555B3	chr14q32.33	230877_at	immunoglobulin heavy constant delta	0	1	1
555B3	chr14q32.33 /// chr16p11.2	211868_x_at	immunoglobulin heavy constant alpha 1 /// immunoglobulin heavy constant gamma 1 (G1m marker) /// immunoglobulin heavy constant gamma 3 (G3m marker) /// similar to Ig heavy chain V-III region VH26 precursor /// myosin-reactive immunoglobulin heavy chain va	0	1	1
555B3	chr14q32.33	213512_at	chromosome 14 open reading frame 79	0	1	1
555B3	chr14q32.32-q32.33	240915_at	immunoglobulin heavy variable 1-69	0	1	1
555B3	chr14q32.33	206478_at	KIAA0125	0	1	1
555B3	chr14q32	209784_s_at	jagged 2	0	1	1
555B3	chr14q32.33	1558438_a_at	Immunoglobulin heavy constant gamma 1 (G1m marker)	0	1	1
555B3	chr14q32.33	209374_s_at	immunoglobulin heavy constant mu	0	1	1
555B3	chr14q	203754_s_at	BRF1 homolog, subunit of RNA polymerase III transcription initiation factor IIIB (S. cerevisiae)	0	1	1
555B3	chr14q32.3	208978_at	cysteine-rich protein 2	1	1	1
555B3	chr14q32.33	234477_at	immunoglobulin heavy constant alpha 1 /// similar to Ig heavy chain V-II region SESS precursor	0	1	1
555B3	chr14q32.33	211430_s_at	immunoglobulin heavy locus /// immunoglobulin heavy constant gamma 1 (G1m marker) /// immunoglobulin heavy constant gamma 2 (G2m marker) /// immunoglobulin heavy constant gamma 3 (G3m marker) /// immunoglobulin heavy constant mu /// anti-RhD monoclonal T1	0	1	1
555B3	chr14q32.33	205081_at	cysteine-rich protein 1 (intestinal)	0	1	1
555B3	chr14q32.33	231910_at	Nudix (nucleoside diphosphate linked moiety X)-type motif 14	0	1	1
555B3	chr14q32.33	217169_at	immunoglobulin heavy constant alpha 1 /// similar to Ig heavy chain V-III region VH26 precursor	0	1	1

555B3	chr14q32.33	216706_x_at	immunoglobulin heavy constant alpha 1 /// immunoglobulin heavy constant delta /// immunoglobulin heavy constant gamma 1 (G1m marker) /// immunoglobulin heavy constant mu	0	1	1
555B3	chr14q32.33	216542_x_at	immunoglobulin heavy constant alpha 1 /// immunoglobulin heavy constant gamma 1 (G1m marker)	0	1	1
555B3	chr14q32.33 /// chr14q32.32-q32.33	216541_x_at	immunoglobulin heavy constant alpha 1 /// immunoglobulin heavy constant gamma 1 (G1m marker) /// immunoglobulin heavy constant gamma 3 (G3m marker) /// immunoglobulin heavy variable 1-69	0	1	1
555B3	chr14q32.33	217022_s_at	immunoglobulin heavy constant alpha 1 /// immunoglobulin heavy constant alpha 2 (A2m marker)	0	1	1
555B3	chr14q32.33	1564499_at	chromosome 14 open reading frame 81	0	1	1
555B3	chr14q32.3	221140_s_at	G protein-coupled receptor 132	0	1	1
555B3	chr14q32.33	220377_at	family with sequence similarity 30, member A	0	1	1
555B3	chr14q32.33	222914_s_at	transmembrane protein 121	1	1	1
555B3	chr14q32.33	217198_x_at	immunoglobulin heavy locus /// immunoglobulin heavy constant delta /// immunoglobulin heavy constant gamma 1 (G1m marker) /// anti-RhD monoclonal T125 gamma1 heavy chain	0	1	1
555B3	chr14q32.33	218399_s_at	cell division cycle associated 4	0	1	1
555B3	chr14q32	225389_at	BTB (POZ) domain containing 6	0	1	1
555B3	chr14q32.33 /// chr16p11.2	217369_at	immunoglobulin heavy constant gamma 1 (G1m marker) /// similar to Ig heavy chain V-III region VH26 precursor /// anti-RhD monoclonal T125 gamma1 heavy chain	0	1	1
555B3	chr14q32.33	217360_x_at	immunoglobulin heavy constant alpha 1 /// immunoglobulin heavy constant gamma 1 (G1m marker) /// immunoglobulin heavy constant gamma 3 (G3m marker) /// similar to Ig heavy chain V-III region VH26 precursor /// similar to Ig heavy chain V-III region VH26 p	0	1	1
555B3	chr14q32.33	1558378_a_at	chromosome 14 open reading frame 78	0	1	1
555B3	chr14q32.33	217217_at	Variable region of IgA (VH4 family) /// Immunoglobulin heavy constant gamma 1 (G1m marker)	1	1	1
555B3	chr14q32.33	1558581_at	Hypothetical protein LOC647310	0	1	1
555B3	chr14q32.33	215949_x_at	immunoglobulin heavy constant mu /// similar to Ig heavy chain V-III region VH26 precursor /// similar to Ig heavy chain V-III region VH26 precursor	1	1	1
555B3	chr14q32.33	215721_at	immunoglobulin heavy constant gamma 1 (G1m marker) /// similar to Ig heavy chain V region 102 precursor	0	1	1
1401B3	chr15q11.2 /// chr14q11.2	1560734_at	olfactory receptor, family 4, subfamily N, member 4 /// olfactory receptor, family 4, subfamily Q, member 3	0	1	1
1401B3	chr15q11.2	1564856_s_at	Olfactory receptor, family 4, subfamily N, member 4	0	1	1

334T2	chr15q11.2	1552696_at	non imprinted in Prader-Willi/Angelman syndrome 1	1	1	1
1229T2	chr15q11.2	1564855_at	Olfactory receptor, family 4, subfamily M, member 2	0	1	1
334T2	chr15q11.2	227967_at	tubulin, gamma complex associated protein 5	1	1	1
334T2	chr15q11	208923_at	cytoplasmic FMR1 interacting protein 1	0	1	1
334T2	chr15q11.2 /// chr15q13.1	217520_x_at	hypothetical protein LOC283683 /// similar to programmed cell death 6 interacting protein	0	1	1
334T2	chr15q11.2	212129_at	non imprinted in Prader-Willi/Angelman syndrome 2	0	1	1
334T2	chr15q13.1 /// chr15q11.2	222149_x_at	golgi autoantigen, golgin subfamily a, 8G /// golgi autoantigen, golgin subfamily a, 8D /// golgi autoantigen, golgin subfamily a, 8E /// golgi autoantigen, golgin subfamily a, 8C /// golgi autoantigen, golgin subfamily a, 8F	0	1	1
497B3	chr15q14	1561405_s_at	cation channel, sperm associated 2	0	1	1
497B3	chr15q15.2	1561705_at	Tau tubulin kinase 2	0	1	1
497B3	chr15q14	1552680_a_at	cancer susceptibility candidate 5	0	1	1
497B3	chr15q15.1	1563079_at	Hypothetical protein LOC645022	0	1	1
497B3	chr15q15.1	1557864_x_at	Phospholipase A2, group IVE	0	1	1
497B3	chr15q15.1	1562163_at	Nucleolar and spindle associated protein 1	0	1	1
497B3	chr15q15-q21	203050_at	tumor protein p53 binding protein, 1	0	1	1
497B3	chr15q15.1-q21.1	1558273_a_at	TYRO3 protein tyrosine kinase	0	1	1
497B3	chr15q15.1	1566209_at	Mitogen activated protein kinase binding protein 1	0	1	1
497B3	chr15q15.1	202826_at	serine peptidase inhibitor, Kunitz type 1	0	1	1
497B3	chr15q15	202712_s_at	creatine kinase, mitochondrial 1B /// creatine kinase, mitochondrial 1A	1	1	1
497B3	chr15q15.1	1560814_a_at	coiled-coil domain containing 32	0	1	1
497B3	chr15q14	1561320_at	P21(CDKN1A)-activated kinase 6	0	1	1
497B3	chr15q15.3	1561306_s_at	stereocilin /// similar to stereocilin /// similar to stereocilin	0	1	1
497B3	chr15q24.2	1566934_at	TYRO3P protein tyrosine kinase pseudogene	0	1	1
497B3	chr15q14-q15	1552276_a_at	vacuolar protein sorting protein 18	0	1	1
497B3	chr15q15.3	1555255_a_at	Histidine acid phosphatase domain containing 2A /// hypothetical protein LOC649951	0	1	1
497B3	chr15q15.3	1569470_a_at	FERM domain containing 5	0	1	1
497B3	chr15q15.1	1569065_s_at	hypothetical protein LOC643338	0	1	1
497B3	chr15q11.1	1556607_at	EH-domain containing 4	0	1	1
497B3	chr15q22	200007_at	signal recognition particle 14kDa (homologous Alu RNA binding protein) /// signal recognition particle 14kDa (homologous Alu RNA binding protein)	0	1	1
497B3	chr15q15.1	1558947_at	Opa interacting protein 5	0	1	1
497B3	chr15q14	1560081_at	hypothetical protein LOC90408	0	1	1
497B3	chr15q14	1564640_at	MAX gene associated	0	1	1
497B3	chr15q15.1	1553907_a_at	exonuclease 3'-5' domain-like 1	0	1	1

497B3	chr15q15-q21.1 /// chr9p11.2	1554741_s_at	fibroblast growth factor 7 (keratinocyte growth factor) /// keratinocyte growth factor-like protein 1 /// similar to Keratinocyte growth factor precursor (KGF) (Fibroblast growth factor 7) (FGF-7) (HBGF-7) /// similar to Keratinocyte growth factor precurs	0	1	1
497B3	chr15q15.1	1554914_at	phospholipase A2, group IVD (cytosolic)	0	1	1
497B3	chr15q14	1557628_s_at	hypothetical protein LOC283745	0	1	1
497B3	chr15q15.1	220082_at	protein phosphatase 1, regulatory (inhibitor) subunit 14D	0	1	1
497B3	chr15q15.1	219861_at	DnaJ (Hsp40) homolog, subfamily C, member 17	0	1	1
497B3	chr15q15.1	231431_s_at	hypothetical LOC388114 /// hypothetical protein LOC649557	0	1	1
497B3	chr15q15.3	236018_at	Leucine carboxyl methyltransferase 2	0	1	1
497B3	chr15q15.1	220071_x_at	centrosomal protein 27kDa	1	1	1
497B3	chr15q15.2	231080_at	Congenital dyserythropoietic anemia, type I	0	1	1
497B3	chr15q15.1	230843_at	Family with sequence similarity 82, member C	1	1	1
497B3	chr15q15.1	230832_at	Rtf1, Paf1/RNA polymerase II complex component, homolog (S. cerevisiae)	0	1	1
497B3	chr15q21	220067_at	spectrin, beta, non-erythrocytic 5	0	1	1
497B3	chr15q15.3	217756_x_at	small EDRK-rich factor 2	0	1	1
497B3	chr15q15.1	217781_s_at	zinc finger protein 106 homolog (mouse)	0	1	1
497B3	chr15q15	235452_at	Creatine kinase, mitochondrial 1A	0	1	1
497B3	chr15q15.3	219518_s_at	elongation factor RNA polymerase II-like 3	0	1	1
497B3	chr15q15.3	218680_x_at	Huntingtin interacting protein K	0	1	1
497B3	chr15q15.2	235340_at	glucosidase, alpha; neutral C	0	1	1
497B3	chr15q15.2	218776_s_at	transmembrane protein 62	0	1	1
497B3	chr15q11.2-q21.3	219095_at	phospholipase A2, group IVB (cytosolic)	0	1	1
497B3	chr15q15.1	219270_at	ChaC, cation transport regulator-like 1 (E. coli)	0	1	1
497B3	chr15q15.1	229579_s_at	dispatched homolog 2 (Drosophila)	0	1	1
497B3	chr15q15.1	218441_s_at	RNA polymerase II associated protein 1	0	1	1
497B3	chr15q15	227033_at	protein disulfide isomerase family A, member 3	1	1	1
497B3	chr15q15.1	230485_at	hypothetical protein LOC644844	0	1	1
497B3	chr15q15.1	225843_at	zinc finger, FYVE domain containing 19	0	1	1
497B3	chr15q15.1	226314_at	dermatan 4 sulfotransferase 1	0	1	1
497B3	chr15q14	226530_at	Bcl2 modifying factor	0	1	1
497B3	chr15q15.3	226562_at	zinc finger protein 690	0	1	1
497B3	chr15q13	226921_at	ubiquitin protein ligase E3 component n-recognin 1	0	1	1
497B3	chr15q15.1	225357_s_at	INO80 complex homolog 1 (S. cerevisiae)	0	1	1
497B3	chr15q15.2	227108_at	START domain containing 9	0	1	1
497B3	chr15q15.1	227272_at	FLJ43339 protein	0	1	1
497B3	chr15q14-q15.1	227846_at	G protein-coupled receptor 176	0	1	1

497B3	chr15q14	223525_at	delta-like 4 (Drosophila)	0	1	1
497B3	chr15q14 /// chr15q15.3	217588_at	cation channel, sperm associated 2 /// cation channel, sperm associated 2 pseudogene	0	1	1
497B3	chr15q15.1	203051_at	bromo adjacent homology domain containing 1	0	1	1
497B3	chr15q15.1	225567_at	Hypothetical LOC388114	0	1	1
497B3	chr15q13.3	221940_at	RNA pseudouridylate synthase domain containing 2	0	1	1
497B3	chr15q15.1	223771_at	Transmembrane protein 87A	0	1	1
497B3	chr15q15.1	229232_at	leucine rich repeat containing 57	0	1	1
497B3	chr15q15.1	225164_s_at	eukaryotic translation initiation factor 2 alpha kinase 4	0	1	1
497B3	chr15q15.1	225300_at	chromosome 15 open reading frame 23	0	1	1
497B3	chr15q14-q15	225311_at	isovaleryl Coenzyme A dehydrogenase	0	1	1
497B3	chr15q14-q15	223084_s_at	cyclin D-type binding-protein 1	1	1	1
497B3	chr15q15.1-q21.1	210944_s_at	calpain 3, (p94)	0	1	1
497B3	chr15q15.1-q21.1	207106_s_at	leukocyte tyrosine kinase	1	1	1
497B3	chr15q21	207121_s_at	mitogen-activated protein kinase 6	0	1	1
497B3	chr15q15.2	207911_s_at	transglutaminase 5	1	1	1
497B3	chr15q13.3	207993_s_at	calcium binding protein P22	1	1	1
497B3	chr15q13.3	241990_at	ras homolog gene family, member V	0	1	1
497B3	chr15q15-q21.1	205782_at	fibroblast growth factor 7 (keratinocyte growth factor)	1	1	1
497B3	chr15q15.3	240048_at	stereocilin	0	1	1
497B3	chr15q15.1	209130_at	synaptosomal-associated protein, 23kDa	0	1	1
497B3	chr15q15	210388_at	phospholipase C, beta 2	0	1	1
497B3	chr15q15.1	235958_at	Phospholipase A2, group IVF	0	1	1
497B3	chr15q15-q21	240274_at	Erythrocyte membrane protein band 4.2	0	1	1
497B3	chr15q13-qter	203151_at	microtubule-associated protein 1A	0	1	1
497B3	chr15q15-q21	203406_at	microfibrillar-associated protein 1	1	1	1
497B3	chr15q15	203755_at	BUB1 budding uninhibited by benzimidazoles 1 homolog beta (yeast)	0	1	1
497B3	chr15q14-q21	205874_at	inositol 1,4,5-trisphosphate 3-kinase A	0	1	1
497B3	chr15q11.2-q21.3	204125_at	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, assembly factor 1	0	1	1
497B3	chr15q15	211337_s_at	gamma tubulin ring complex protein (76p gene)	1	1	1
497B3	chr15q15.2	243949_at	Similar to kinesin-like motor protein C20orf23	0	1	1
497B3	chr15q15	204867_at	GTP cyclohydrolase I feedback regulator	0	1	1
497B3	chr15q15.1	205023_at	RAD51 homolog (RecA homolog, E. coli) (S. cerevisiae)	0	1	1
497B3	chr15q15.3	205519_at	WD repeat domain 76	0	1	1
497B3	chr15q15.3	238424_at	adenosine deaminase-like	0	1	1
497B3	chr15q15-q21	210746_s_at	erythrocyte membrane protein band 4.2 /// erythrocyte membrane protein band 4.2	1	1	1
497B3	chr15q15.1	239470_at	hypothetical protein LOC644809	0	1	1

497B3	chr15q15.2	237288_at	transglutaminase 7	0	1	1
497B3	chr15q15.1	212156_at	vacuolar protein sorting 39 (yeast)	0	1	1
497B3	chr15q24.1	213455_at	hypothetical LOC283677	0	1	1
956T1	chr15q	213012_at	neural precursor cell expressed, developmentally down-regulated 4	1	1	0
1014T2	chr15q26.3	1554456_a_at	lines homolog 1 (Drosophila)	0	0	0
956T1	chr16p13.3	1552349_a_at	protease, serine, 33	0	1	0
956T1	chr16p12.3	200085_s_at	transcription elongation factor B (SIII), polypeptide 2 (18kDa, elongin B) /// transcription elongation factor B (SIII), polypeptide 2 (18kDa, elongin B)	0	1	0
956T1	chr16p13.3	220051_at	protease, serine, 21 (testisin)	0	1	0
956T1	chr16p13.3	205847_at	protease, serine, 22	0	1	0
956T1	chr16p13.3	216629_at	Serine/arginine repetitive matrix 2	0	1	0
956T1	chr16p13.3	222170_at	Hypothetical LOC440334	0	1	0
956T1	chr16p13.3	228058_at	similar to common salivary protein 1	0	1	0
956T1	chr16p12.3	213877_x_at	transcription elongation factor B (SIII), polypeptide 2 (18kDa, elongin B)	0	1	0
956T1	chr16p13.3	226079_at	hypothetical protein BC014089	0	1	0
1052T1	chr16p11.2	239993_at	similar to protein phosphatase 2A 48 kDa regulatory subunit isoform 1; serine/threonine protein phosphatase 2A, 48kDa regulatory subunit; PP2A, subunit B, PR48 isoform; PP2A B subunit PR48; NY-REN-8 antigen /// similar to protein phosphatase 2, regulatory	0	1	1
588T1	chr16p11.2	1553914_at	hypothetical protein MGC34800	0	1	1
1052T1	chr16q13	216336_x_at	metallothionein 1M	1	1	1
1052T1	chr16p11.2	216193_at	hect domain and RLD 2 pseudogene	1	1	1
1052T1	chr16p13 /// chr16p11.2	220167_s_at	TP53TG3 protein /// similar to TP53TG3 protein /// similar to TP53TG3 protein /// similar to TP53TG3 protein	0	1	1
588T1	chr16p11.1	1561518_at	hypothetical protein LOC283914	0	1	1
588T1	chr16p11.2	217384_x_at	similar to Ig heavy chain V-III region VH26 precursor /// similar to Ig heavy chain V-III region VH26 precursor /// similar to Ig heavy chain V-III region VH26 precursor	0	1	1
1052T1	chr16q24.2-q24.3	204726_at	cadherin 13, H-cadherin (heart)	1	1	0
1469B1	chr17q21.31	237962_x_at	KIAA1267	0	1	1
1469B1	chr17q21.32 /// chr17q21.31	1555794_at	ADP-ribosylation factor-like 17 pseudogene 1 /// ADP-ribosylation factor-like protein	0	1	1
1469B1	chr17q24.3	230056_at	fetal Alzheimer antigen	0	1	1
1469B1	chr17q21.31	230388_s_at	hypothetical protein LOC644246 /// hypothetical protein LOC649063	0	1	1
1469B1	chr17q21.31	229857_s_at	Hypothetical protein LOC644246	0	1	1
810T1	chr18q21.1	202002_at	acetyl-Coenzyme A acyltransferase 2 (mitochondrial 3-oxoacyl-Coenzyme A thiolase)	0	1	0

1052T1	chr19p13.2	1553276_at	zinc finger protein 560	0	1	0
1014T2	chr19p13.11	243482_at	Epidermal growth factor receptor pathway substrate 15-like 1	1	1	0
1014T2	chr19p13.11	1552421_a_at	calreticulin 3	0	1	0
464T3	chr19q13.42	213402_at	hypothetical protein LOC126208	0	1	0
1376T2	chr19q13.43	217593_at	zinc finger protein 447	0	1	0
1466T2	chr19qter	204937_s_at	zinc finger protein 274	1	1	0
1376T2	chr19p12	1569637_at	zinc finger protein 100	0	1	0
1466T2	chr19q13.43	218735_s_at	zinc finger protein 544	0	1	0
1376T2	chr19q13.43	219765_at	zinc finger protein 329	1	1	0
1376T2	chr19q13.43	218735_s_at		0	1	0
1401B3	chr20p13	231905_at	chromosome 20 open reading frame 96	0	1	1
1401B3	chr20p13	233160_at	defensin, beta 129	0	1	1
699T2a	chr20p12.3	224826_at	hypothetical protein KIAA1434	1	1	0
699T2a	chr20p13	233630_at	CDP-diacylglycerol synthase (phosphatidate cytidylyltransferase) 2	1	1	0
699T2a	chr20p12.3	229992_at	hypothetical protein LOC149837	0	1	0
1166B3	chr20p11.2	219913_s_at	Crn, crooked neck-like 1 (Drosophila)	0	1	0
1166B3	chr20p11.23	233389_at	chromosome 20 open reading frame 26	0	0	0
597B3	chr21p11	220205_at	transmembrane phosphatase with tensin homology	0	1	1
862T1	chr21q22.1-q22.2|21q22.11	200818_at	ATP synthase, H+ transporting, mitochondrial F1 complex, O subunit (oligomycin sensitivity conferring protein)	1	0	0
956T1	chr21q22.2|21q22.13	1557900_at	single-minded homolog 2 (Drosophila)	0	1	0
1399B3	chr22q11.1	220508_at	T-complex protein 1	0	1	1
1399B3	chr22q11.2	233469_at	TPTE pseudogene	0	1	1

[a] Locations of probe sets from the HG-U133 Plus 2 GeneChip were compared to CNV regions. This gene list has been filtered to include only unique genes.

[b] 1, probe sets differentially expressed in AD vs. controls

[c] 1, probe sets located in coding regions

[d] 1, probe sets located within known CNV regions published in the Database for Genomic Variants

Chapter V. Conclusions

A. Summary

Advances in technology have led to a dramatic increase in the ability to identify genes involved in susceptibility to common complex traits and to elucidate underlying biological processes [8, 27, 146]. High throughput, whole genome gene profiling and whole genome genotyping assays are at the heart of these accomplishments. Combining these techniques can reveal unpredicted and surprising molecular mechanisms for phenotypic outcomes. However, there are challenges in using them for biological discovery.   Novel sophisticated techniques require novel analytic methods.   The difficulty is to create robust analytic methods capable of distinguishing biological meaningful signals of varying effect size from a background of complex variability and with the additional statistical considerations of small sample sizes and multiple testing issues.

This research focused on utilizing these high throughput methods as alternative strategies for gene mapping in Alzheimer's disease, a common multifactorial disease involving multiple cellular processes and environmental influences. Alzheimer's disease is the most frequent cause of cognitive decline in the elderly and afflicts approximately 26 million individuals worldwide [59]. We proposed that using gene expression results to inform decisions of candidate gene association studies for AD would improve our ability to choose rational gene candidates.

Our first step was to conduct a gene profiling approach to dissecting the complex phenotypes involved in age-related cognitive decline. Comparison of our results with two previously published studies using a comparable microarray platform revealed common pathways underlying cognitive decline in three different brain tissues. Novel genes in pathways previously recognized as crucial to healthy brain aging have been identified. Sweeping transcriptional differences associated with cognitive decline implicated genes involved in transcriptional regulation, energy pathways, ion homeostasis dysregulation, apoptosis, and synaptic activity. In addition, our results reveal significant up-regulation of actin-related processes and down-regulation of translation, RNA processing and localization, and vesicle mediated transport. This study identifies 873 candidate genes, located in linkage regions, which had not been previously implicated in cognitive decline. Dysregulated genes that are both involved in known AD critical pathways and located in linkage/association regions represent potential candidates for gene association studies.

Recent studies into the genetics of gene expression [122-124] have shown that transcriptional response to environmental and cellular perturbations may vary from person to person based on specific genetic sequence. Advances in genotyping technology have led to a dramatic drop in cost per SNP and allowed us to test the extent to which our gene expression profiling results were affected by copy number variation by performing whole genome SNP genotyping in the same subjects. Low level analysis of two different genotyping algorithms demonstrated the importance of determining the sensitivity of an algorithm to the physicochemical properties of the probes on the array. We found that genotyping call rates across the Affymetrix 100K Mapping Set arrays can be improved by choosing an algorithm such as BRLMM that corrects for the length of the PCR fragment hybridized to the chip and the GC content of the probe itself. The improvement in data quality increases power to discover disease susceptibility genes.

Although advances in genomic technology have begun to revolutionize human biology and genetics medicine, they have been focused largely on the nature and pattern of SNPs within the human genome. Recent studies have shown the importance of larger DNA sequence polymorphisms up to 2Mb in size. There has been considerable progress in understanding the common patterns of SNPs, but the extent and impact of structural variation in normal individuals is still being determined. We tested the effect of CNV on genotyping call rate and on gene expression patterns. Several issues were identified in this approach: 1) the importance of identifying the features requiring normalization across the samples, 2) the impact of reference data set choice has on the detection and estimation of copy numbers differences, 3) there is an increase in NoCall rate in CNV regions due to the polymorphic nature of SNPs and 2) there is a decreased density of SNPs in duplicated regions of the genome which limit our power to detect CNV regions across the genome. This relates to the studies showing that genetic complexity of structural genetic variation decreases the linkage disequilibrium (LD) in surrounding areas of the genome [8, 147]. Because of this, the SNPs captured on the manufactured genotyping arrays are not necessarily effective proxies for nearby CNV and therefore, it is important to investigate the association of CNV with a complex trait separately from association with SNP and disease.

B. Future Directions

Whole genome association studies have an advantage over candidate gene studies in that novel genes and biological pathways not previously thought to be involved can be discovered. Both whole genome and candidate gene studies will be needed to replicate and confirm candidates found in this study. Effects of the massive transcriptional response to cognitive decline reflect the complex interplay between transcriptional regulation and metabolism in response to the environment. A major new emphasis in genomic medicine will be untangling what the specific genetic variants do within each environmental and biological process.   Multiple molecular changes can result from environmental stressors, but not all of these changes are linked to increased disease risk. In particular, a late onset disease has its beginnings years earlier and has shaped downstream cellular effects over time.   In addition, long term environmental contributions to chronic disease are likely different from acute environmental effects.

Additional studies are needed to address these issues and refine the role of gene expression and DNA structural variability in complex traits. An important issue in this study is that the brain tissue used is comprised of mixed cell types from end stage AD. Experiments from early stages of disease would be useful to identify genes in the begining stages of the neurodegenerative process. Likewise, using single cells in place of mixed tissues, would identify cell specific gene expression changes. Replication of these experiments or combination with other metadata is necessary to see global transcriptional reprogramming due to cognitive decline versus stochastic changes due to natural plasticity of the regulatory system [148]. Investigations into splice variants and tissue specificity of these variants will be important to elucidate the role of specific isoforms in the disease process. Whole genome experimental designs are particularly useful for discovery of regulatory regions of the genome ignored by candidate gene studies. Epistatic interactions among SNPs within and among genes, CNVs, SNPs x CNVs and environmental interactions could have either synergistic or compensatory effects depending on the genotypes. This is a particularly challenging problem due to the high dimensional nature of the technology and the lack of a priori hypotheses[149].

The integration of expression and genomic sequence variation will lead to a better understanding of diseases and designing improved treatments. Fine tuning the SNP array experiments in terms of staging of disease and single cell types as discussed earlier would provide the same deeper understanding of the disease process.

The newer platform chips, Affymetrix 5.0 Assay and 6.0 Assay, are considerably different with respect to density and probes on the array. Determining which probes are underperforming across multiple samples will allow the SNPs affecting whole genome association studies to be identified and removed. SNPs underperforming across multiple populations may suggest structural differences in populations. Also, the use of chips with higher density SNPs will allow finer mapping of copy number breakpoints. As genotyping platforms and algorithms continually evolve, there is an ongoing need for testing and evaluation to determine under what conditions these algorithms improve the sensitivity and specificity of the genotype calls.

The etiology of most common complex traits has multiple genetic and environmental factors. Also, the chronic effects of late onset diseases shape downstream cellular effects over time but are not subject to evolutionary constraints that select for particular mutations. This may make the identification of genes involved in these phenotypes more difficult because of possible increase in the number of genes involved and the possibility that CNV have rearranged on multiple haplotype backgrounds (ie., decreased linkage disequilibrium). Such genes will likely have smaller effect sizes requiring much larger sample sizes in order to identify them. To make such studies practical, consortium based approaches such as the Wellcome Trust Case Control Consortium (WTCCC) which has recently identified 24 genetic risk factors for seven common complex diseases [30] will be needed. Finally, given the complex nature of these phenotypes, identification and characterization of confounding factors is a key component in analyzing population data. Movements such as the recent NIH call (RFA-HG-07-005) to set up a consortium of sites to perform whole genome association studies on tissue repositories linked to electronic medical records will be an important resource for investigators studying the genetics of common complex traits, allowing critical annotation to help define patient phenotypes.

Literature

1. Chakravarti, A. and P. Little, *Nature, nurture and human disease.* Nature, 2003. **421**(6921): p. 412-4.
2. Altshuler, D. and A.G. Clark, *Genetics. Harvesting medical information from the human family tree.* Science, 2005. **307**(5712): p. 1052-3.
3. Altshuler, D. and M. Daly, *Guilt beyond a reasonable doubt.* Nat Genet, 2007. **39**(7): p. 813-5.
4. Ardlie, K.G., L. Kruglyak, and M. Seielstad, *Patterns of linkage disequilibrium in the human genome.* Nat Rev Genet, 2002. **3**(4): p. 299-309.
5. Risch, N. and K. Merikangas, *The future of genetic studies of complex human diseases.* Science, 1996. **273**(5281): p. 1516-7.
6. Carlson, C.S., et al., *Mapping complex disease loci in whole-genome association studies.* Nature, 2004. **429**(6990): p. 446-52.
7. Glazier, A.M., J.H. Nadeau, and T.J. Aitman, *Finding genes that underlie complex traits.* Science, 2002. **298**(5602): p. 2345-9.
8. Redon, R., et al., *Global variation in copy number in the human genome.* Nature, 2006. **444**(7118): p. 444-54.
9. Sharp, A.J., et al., *Segmental duplications and copy-number variation in the human genome.* Am J Hum Genet, 2005. **77**(1): p. 78-88.
10. Wong, K.K., et al., *A comprehensive analysis of common copy-number variations in the human genome.* Am J Hum Genet, 2007. **80**(1): p. 91-104.
11. Hinds, D.A., et al., *Common deletions and SNPs are in linkage disequilibrium in the human genome.* Nat Genet, 2006. **38**(1): p. 82-5.

12. Iafrate, A.J., et al., *Detection of large-scale variation in the human genome.* Nat Genet, 2004. **36**(9): p. 949-51.

13. Conrad, D.F., et al., *A high-resolution survey of deletion polymorphism in the human genome.* Nat Genet, 2006. **38**(1): p. 75-81.

14. McCarroll, S.A., et al., *Common deletion polymorphisms in the human genome.* Nat Genet, 2006. **38**(1): p. 86-92.

15. Sebat, J., et al., *Large-scale copy number polymorphism in the human genome.* Science, 2004. **305**(5683): p. 525-8.

16. Schmitt, F.A., et al., *"Preclinical" AD revisited: neuropathology of cognitively normal older adults.* Neurology, 2000. **55**(3): p. 370-6.

17. Katzman, R., et al., *Clinical, pathological, and neurochemical changes in dementia: a subgroup with preserved mental status and numerous neocortical plaques.* Ann Neurol, 1988. **23**(2): p. 138-44.

18. Lee, J.H., *Genetic evidence for cognitive reserve: variations in memory and related cognitive functions.* J Clin Exp Neuropsychol, 2003. **25**(5): p. 594-613.

19. Scarmeas, N. and Y. Stern, *Cognitive reserve: implications for diagnosis and prevention of Alzheimer's disease.* Curr Neurol Neurosci Rep, 2004. **4**(5): p. 374-80.

20. Singleton, A., A. Myers, and J. Hardy, *The law of mass action applied to neurodegenerative disease: a hypothesis concerning the etiology and pathogenesis of complex diseases.* Hum Mol Genet, 2004. **13 Spec No 1**: p. R123-6.

21. Farrall, M., *Quantitative genetic variation: a post-modern view.* Hum Mol Genet, 2004. **13 Spec No 1**: p. R1-7.

22. Jais, P.H., *How frequent is altered gene expression among susceptibility genes to human complex disorders?* Genet Med, 2005. **7**(2): p. 83-96.

23. Yan, H., et al., *Allelic variation in human gene expression.* Science, 2002. **297**(5584): p. 1143.

24. Cheung, V.G., et al., *Natural variation in human gene expression assessed in lymphoblastoid cells.* Nat Genet, 2003. **33**(3): p. 422-5.

25. Morley, M., et al., *Genetic analysis of genome-wide variation in human gene expression.* Nature, 2004. **430**(7001): p. 743-7.

26. Cutler, D.J., et al., *High-throughput variation detection and genotyping using microarrays.* Genome Res, 2001. **11**(11): p. 1913-25.

27. Kennedy, G.C., et al., *Large-scale genotyping of complex DNA.* Nat Biotechnol, 2003. **21**(10): p. 1233-7.

28. Roses, A.D., et al., *Complex disease-associated pharmacogenetics: drug efficacy, drug safety, and confirmation of a pathogenetic hypothesis (Alzheimer's disease).* Pharmacogenomics J, 2006.

29. Kerb, R., *Implications of genetic polymorphisms in drug transporters for pharmacotherapy.* Cancer Lett, 2006. **234**(1): p. 4-33.

30. *Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.* Nature, 2007. **447**(7145): p. 661-78.

31. Irizarry, R.A., et al., *Summaries of Affymetrix GeneChip probe level data.* Nucleic Acids Res, 2003. **31**(4): p. e15.

32. Li, C. and W. Hung Wong, *Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application.* Genome Biol, 2001. **2**(8): p. RESEARCH0032.

33. Huber, W., et al., *Variance stabilization applied to microarray data calibration and to the quantification of differential expression.* Bioinformatics, 2002. **18 Suppl 1**: p. S96-104.

34. Di, X., et al., *Dynamic model based algorithms for screening and genotyping over 100 K SNPs on oligonucleotide microarrays.* Bioinformatics, 2005. **21**(9): p. 1958-63.

35. Rabbee, N. and T.P. Speed, *A genotype calling algorithm for affymetrix SNP arrays.* Bioinformatics, 2006. **22**(1): p. 7-12.

36. Monks, S.A., et al., *Genetic inheritance of gene expression in human cell lines.* Am J Hum Genet, 2004. **75**(6): p. 1094-105.

37. Schadt, E.E., et al., *Genetics of gene expression surveyed in maize, mouse and man.* Nature, 2003. **422**(6929): p. 297-302.

38. Cheung, V.G., et al., *Genetics of quantitative variation in human gene expression.* Cold Spring Harb Symp Quant Biol, 2003. **68**: p. 403-7.

39. Buckland, P.R., *Polymorphically duplicated genes: their relevance to phenotypic variation in humans.* Ann Med, 2003. **35**(5): p. 308-15.

40. Carter, N.P., *As normal as normal can be?* Nat Genet, 2004. **36**(9): p. 931-2.

41. Check, E., *Human genome: patchwork people.* Nature, 2005. **437**(7062): p. 1084-6.

42. Hegele, R.A., *Copy-number variations add a new layer of complexity in the human genome.* Cmaj, 2007. **176**(4): p. 441-2.

43. Sharp, A.J., Z. Cheng, and E.E. Eichler, *Structural variation of the human genome.* Annu Rev Genomics Hum Genet, 2006. **7**: p. 407-42.

44. Feuk, L., A.R. Carson, and S.W. Scherer, *Structural variation in the human genome.* Nat Rev Genet, 2006. **7**(2): p. 85-97.

45. McLeod, H.L. and W.N. Keith, *Variation in topoisomerase I gene copy number as a mechanism for intrinsic drug sensitivity.* Br J Cancer, 1996. **74**(4): p. 508-12.

46. Franchina, M. and P.H. Kay, *Allele-specific variation in the gene copy number of human cytosine 5-methyltransferase.* Hum Hered, 2000. **50**(2): p. 112-7.

47. Townson, J.R., L.F. Barcellos, and R.J. Nibbs, *Gene copy number regulates the production of the human chemokine CCL3-L1.* Eur J Immunol, 2002. **32**(10): p. 3016-26.

48. Linzmeier, R.M. and T. Ganz, *Human defensin gene copy number polymorphisms: comprehensive analysis of independent variation in alpha- and beta-defensin regions at 8p22-p23.* Genomics, 2005. **86**(4): p. 423-30.

49. Bolt, H.M. and R. Thier, *Relevance of the deletion polymorphisms of the glutathione S-transferases GSTT1 and GSTM1 in pharmacology and toxicology.* Curr Drug Metab, 2006. **7**(6): p. 613-28.

50. Cabrejo, L., et al., *Phenotype associated with APP duplication in five families.* Brain, 2006. **129**(Pt 11): p. 2966-76.

51. Shaw-Smith, C., et al., *Microdeletion encompassing MAPT at chromosome 17q21.3 is associated with developmental delay and learning disability*. Nat Genet, 2006. **38**(9): p. 1032-7.

52. Myers, C.L., et al., *Accurate detection of aneuploidies in array CGH and gene expression microarray data.* Bioinformatics, 2004. **20**(18): p. 3533-43.

53. Stransky, N., et al., *Regional copy number-independent deregulation of transcription in cancer.* Nat Genet, 2006. **38**(12): p. 1386-96.

54. Furge, K.A., et al., *Comparison of array-based comparative genomic hybridization with gene expression-based regional expression biases to identify genetic abnormalities in hepatocellular carcinoma.* BMC Genomics, 2005. **6**(1): p. 67.

55. Kloth, J.N., et al., *Combined array-comparative genomic hybridization and single-nucleotide polymorphism-loss of heterozygosity analysis reveals complex genetic alterations in cervical cancer.* BMC Genomics, 2007. **8**: p. 53.

56. Stranger, B.E., et al., *Relative impact of nucleotide and copy number variation on gene expression phenotypes.* Science, 2007. **315**(5813): p. 848-53.

57. Wilmot, B., et al., *Translational gene mapping of cognitive decline.* Neurobiol Aging, 2006.

58. Daruwala, R.S., et al., *A versatile statistical analysis algorithm to detect genome copy number variation.* Proc Natl Acad Sci U S A, 2004. **101**(46): p. 16292-7.

59. Brookmeyer, R., et al., *Forecasting the global burden of Alzheimer’s disease.* Alzheimer's & Dementia: The Journal of the Alzheimer's Association, 2007. **3**(3): p. 186-191.

60. Bertram, L. and R.E. Tanzi, *Alzheimer's disease: one disorder, too many genes?* Human Molecular Genetics, 2004. **13 Spec No 1**: p. R135-41.

61. Braak, H. and E. Braak, *Argyrophilic grain disease: frequency of occurrence in different age categories and neuropathological diagnostic criteria.* J Neural Transm, 1998. **105**(8-9): p. 801-19.

62. Blalock, E.M., et al., *Incipient Alzheimer's disease: microarray correlation analyses reveal major transcriptional and tumor suppressor responses.* Proc Natl Acad Sci U S A, 2004. **101**(7): p. 2173-8.

63. Dunckley, T., et al., *Gene expression correlates of neurofibrillary tangles in Alzheimer's disease.* Neurobiol Aging, 2005.

64. Mirra, S.S., et al., *The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease.* Neurology, 1991. **41**(4): p. 479-86.

65. *Consensus recommendations for the postmortem diagnosis of Alzheimer's disease. The National Institute on Aging, and Reagan Institute Working Group on Diagnostic Criteria for the Neuropathological Assessment of Alzheimer's Disease.* Neurobiol Aging, 1997. **18**(4 Suppl): p. S1-2.

66. Ramakers, C., et al., *Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data.* Neurosci Lett, 2003. **339**(1): p. 62-6.

67. Pfaffl, M.W., *A new mathematical model for relative quantification in real-time RT-PCR.* Nucleic Acids Res, 2001. **29**(9): p. e45.

68. Team, R.D.C., *R: A Language and Environment for Statistical Computing*. 2005, R Foundation for Statistical Computing: Vienna, Austria.

69. Gentleman, R.C., et al., *Bioconductor: open software development for computational biology and bioinformatics.* Genome Biol, 2004. **5**(10): p. R80.

70. Gautier, L., et al., *affy--analysis of Affymetrix GeneChip data at the probe level.* Bioinformatics, 2004. **20**(3): p. 307-15.

71. Dudoit, S., et al., *Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments.* Statistica Sinica, 2002. **12**: p. 111-139.

72. Smyth, G., *Linear models and empirical bayes methods for assessing differential expression in microarray experiments.* Statistical Applications in Genetics and Moliecular Biology, 2004. **3**(1): p. 1-23.

73. Storey, J.D. and R. Tibshirani, *Statistical significance for genomewide studies.* Proc Natl Acad Sci U S A, 2003. **100**(16): p. 9440-5.

74. Scearce, L.M., et al., *Functional genomics of the endocrine pancreas: the pancreas clone set and PancChip, new resources for diabetes research.* Diabetes, 2002. **51**(7): p. 1997-2004.

75. Beissbarth, T. and T.P. Speed, *GOstat: find statistically overrepresented Gene Ontologies within a group of genes.* Bioinformatics, 2004. **20**(9): p. 1464-5.

76. Benjamini, Y. and D. Yekutieli, *The Control of the False Discovery Rate in Multiple Testing under Dependency.* The Annal of Statistics, 2001. **29**(4): p. 1165-1188.

77. Colangelo, V., et al., *Gene expression profiling of 12633 genes in Alzheimer hippocampal CA1: transcription and neurotrophic factor down-regulation and*

up-regulation of apoptotic and pro-inflammatory signaling. Journal of Neuroscience Research, 2002. **70**(3): p. 462-73.

78. Loring, J.F., et al., *A gene expression profile of Alzheimer's disease.* DNA Cell Biol, 2001. **20**(11): p. 683-95.

79. Yao, P.J., et al., *Defects in expression of genes related to synaptic vesicle trafficking in frontal cortex of Alzheimer's disease.* Neurobiol Dis, 2003. **12**(2): p. 97-109.

80. Scheff, S.W., et al., *Hippocampal synaptic loss in early Alzheimer's disease and mild cognitive impairment.* Neurobiol Aging, 2005.

81. Cataldo, A.M., et al., *Endocytic pathway abnormalities precede amyloid beta deposition in sporadic Alzheimer's disease and Down syndrome: differential effects of APOE genotype and presenilin mutations.* Am J Pathol, 2000. **157**(1): p. 277-86.

82. O'Bryan, J.P., R.P. Mohney, and C.E. Oldham, *Mitogenesis and endocytosis: What's at the INTERSECTIoN?* Oncogene, 2001. **20**(44): p. 6300-8.

83. Keating, D.J., C. Chen, and M.A. Pritchard, *Alzheimer's disease and endocytic dysfunction: Clues from the Down syndrome-related proteins, DSCR1 and ITSN1.* Ageing Res Rev, 2006.

84. McPherson, P.S., B.K. Kay, and N.K. Hussain, *Signaling on the endocytic pathway.* Traffic, 2001. **2**(6): p. 375-84.

85. Predescu, S.A., et al., *Intersectin regulates fission and internalization of caveolae in endothelial cells.* Mol Biol Cell, 2003. **14**(12): p. 4997-5010.

86. Pucharcos, C., X. Estivill, and S. de la Luna, *Intersectin 2, a new multimodular protein involved in clathrin-mediated endocytosis.* FEBS Lett, 2000. **478**(1-2): p. 43-51.

87. Tong, X.K., et al., *Intersectin can regulate the Ras/MAP kinase pathway independent of its role in endocytosis.* J Biol Chem, 2000. **275**(38): p. 29894-9.

88. Chyung, J.H. and D.J. Selkoe, *Inhibition of receptor-mediated endocytosis demonstrates generation of amyloid beta-protein at the cell surface.* J Biol Chem, 2003. **278**(51): p. 51035-43.

89. Carey, R.M., et al., *Inhibition of dynamin-dependent endocytosis increases shedding of the amyloid precursor protein ectodomain and reduces generation of amyloid beta protein.* BMC Cell Biol, 2005. **6**: p. 30.

90. Savdie, C., et al., *Cell-type-specific pathways of neurotensin endocytosis.* Cell Tissue Res, 2005: p. 1-17.

91. Mohney, R.P., et al., *Intersectin activates Ras but stimulates transcription through an independent pathway involving JNK.* J Biol Chem, 2003. **278**(47): p. 47038-45.

92. Maekawa, M., et al., *Signaling from Rho to the actin cytoskeleton through protein kinases ROCK and LIM-kinase.* Science, 1999. **285**(5429): p. 895-8.

93. Sun-Wada, G.H., Y. Wada, and M. Futai, *Diverse and essential roles of mammalian vacuolar-type proton pump ATPase: toward the physiological understanding of inside acidic compartments.* Biochim Biophys Acta, 2004. **1658**(1-2): p. 106-14.

94. Bannai, H., et al., *Efficiently finding regulatory elements using correlation with gene expression.* J Bioinform Comput Biol, 2004. **2**(2): p. 273-88.

95. Kanai, Y., N. Dohmae, and N. Hirokawa, *Kinesin transports RNA: isolation and characterization of an RNA-transporting granule.* Neuron, 2004. **43**(4): p. 513-25.

96. Mizutani, A., et al., *SYNCRIP, a cytoplasmic counterpart of heterogeneous nuclear ribonucleoprotein R, interacts with ubiquitous synaptotagmin isoforms.* J Biol Chem, 2000. **275**(13): p. 9823-31.

97. Salehi, A., J.D. Delcroix, and D.F. Swaab, *Alzheimer's disease and NGF signaling.* J Neural Transm, 2004. **111**(3): p. 323-45.

98. Dimakopoulos, A.C., *Protein aggregation in Alzheimer's disease and other neuropathological disorders.* Curr Alzheimer Res, 2005. **2**(1): p. 19-28.

99. Layfield, R., J. Lowe, and L. Bedford, *The ubiquitin-proteasome system and neurodegenerative disorders.* Essays Biochem, 2005. **41**: p. 157-71.

100. Hegde, A.N., *Ubiquitin-proteasome-mediated local protein degradation and synaptic plasticity.* Prog Neurobiol, 2004. **73**(5): p. 311-57.

101. Kita, H., et al., *Modulation of polyglutamine-induced cell death by genes identified by expression profiling.* Hum Mol Genet, 2002. **11**(19): p. 2279-87.

102. Wellmann, S., et al., *Oxygen-regulated expression of the RNA-binding proteins RBM3 and CIRP by a HIF-1-independent mechanism.* J Cell Sci, 2004. **117**(Pt 9): p. 1785-94.

103. Dresios, J., et al., *Cold stress-induced protein Rbm3 binds 60S ribosomal subunits, alters microRNA levels, and enhances global protein synthesis.* Proc Natl Acad Sci U S A, 2005. **102**(6): p. 1865-70.

104. Mattick, J.S. and I.V. Makunin, *Small regulatory RNAs in mammals.* Hum Mol Genet, 2005. **14 Spec No 1**: p. R121-32.

105. Martin, K.C. and K.S. Kosik, *Synaptic tagging -- who's it?* Nat Rev Neurosci, 2002. **3**(10): p. 813-20.

106. Bierer, L.M., et al., *Neocortical neurofibrillary tangles correlate with dementia severity in Alzheimer's disease.* Arch Neurol, 1995. **52**(1): p. 81-8.

107. Stern, Y., *What is cognitive reserve? Theory and research application of the reserve concept.* J Int Neuropsychol Soc, 2002. **8**(3): p. 448-60.

108. Ohki-Hamazaki, H., M. Iwabuchi, and F. Maekawa, *Development and function of bombesin-like peptides and their receptors.* Int J Dev Biol, 2005. **49**(2-3): p. 293-300.

109. Ohki-Hamazaki, H., et al., *Mice lacking bombesin receptor subtype-3 develop metabolic defects and obesity.* Nature, 1997. **390**(6656): p. 165-9.

110. de la Monte, S.M. and J.R. Wands, *Review of insulin and insulin-like growth factor expression, signaling, and malfunction in the central nervous system: relevance to Alzheimer's disease.* J Alzheimers Dis, 2005. **7**(1): p. 45-61.

111. Mosconi, L., *Brain glucose metabolism in the early and specific diagnosis of Alzheimer's disease. FDG-PET studies in MCI and AD.* Eur J Nucl Med Mol Imaging, 2005. **32**(4): p. 486-510.

112. Bernardini, S., et al., *Glutathione S-transferase P1 *C allelic variant increases susceptibility for late-onset Alzheimer disease: association study and relationship with apolipoprotein E epsilon4 allele.* Clin Chem, 2005. **51**(6): p. 944-51.

113. Stroombergen, M.C. and R.H. Waring, *Determination of glutathione S-transferase mu and theta polymorphisms in neurological disease.* Hum Exp Toxicol, 1999. **18**(3): p. 141-5.

114. Hermani, A., et al., *S100A8 and S100A9 activate MAP kinase and NF-kappaB signaling pathways and trigger translocation of RAGE in human prostate cancer cells.* Exp Cell Res, 2006. **312**(2): p. 184-97.

115. Hsu, Y.C. and M.S. Perin, *Human neuronal pentraxin II (NPTX2): conservation, genomic structure, and chromosomal localization.* Genomics, 1995. **28**(2): p. 220-7.

116. Tsui, C.C., et al., *Narp, a novel member of the pentraxin family, promotes neurite outgrowth and is dynamically regulated by neuronal activity*. J Neurosci, 1996. **16**(8): p. 2463-78.

117. Ohtsuka, T., et al., *nRap GEP: a novel neural GDP/GTP exchange protein for rap1 small G protein that interacts with synaptic scaffolding molecule (S-SCAM).* Biochem Biophys Res Commun, 1999. **265**(1): p. 38-44.

118. Marshall, E., *Getting the noise out of gene arrays.* Science, 2004. **306**(5696): p. 630-1.

119. Fathallah-Shaykh, H.M., *Microarrays: applications and pitfalls.* Arch Neurol, 2005. **62**(11): p. 1669-72.

120. Dallas, P.B., et al., *Gene expression levels assessed by oligonucleotide microarray analysis and quantitative real-time RT-PCR -- how well do they correlate?* BMC Genomics, 2005. **6**(1): p. 59.

121. Chanock, S.J., et al., *Replicating genotype-phenotype associations.* Nature, 2007. **447**(7145): p. 655-60.

122. Dermitzakis, E.T. and B.E. Stranger, *Genetic variation in human gene expression.* Mamm Genome, 2006. **17**(6): p. 503-8.

123. Cheung, V.G., et al., *Mapping determinants of human gene expression by regional and genome-wide association.* Nature, 2005. **437**(7063): p. 1365-9.

124. Spielman, R.S., et al., *Common genetic variants account for differences in gene expression among ethnic groups.* Nat Genet, 2007. **39**(2): p. 226-31.

125. Couzin, J. and J. Kaiser, *Genome-wide association. Closing the net on common disease genes.* Science, 2007. **316**(5826): p. 820-2.

126. Team, R.D.C., *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. 2006: Vienna, Austria.

127. Venables, W.N. and B.D. Ripley, *Modern Applied Statistics with S.* Fourth Edition ed. 2002, New York: Springer.

128. Harrell, F.E., Jr., *Hmisc*. 2006.

129. Warnes, G. and F. Leisch, *The Genetics Package*. 2005.

130. Benjamini, Y. and Y. Hochberg, *Controlling the false discovery rate: a practical and powerful apporach to multiple testing.* J Royal Stat Soc, 1995. **Ser B 57**: p. 289-300.

131. Nannya, Y., et al., *A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays.* Cancer Res, 2005. **65**(14): p. 6071-9.

132. Komura, D., et al., *Noise reduction from genotyping microarrays using probe level information.* In Silico Biol, 2006. **6**(1-2): p. 79-92.

133. Ishikawa, S., et al., *Allelic dosage analysis with genotyping microarrays.* Biochem Biophys Res Commun, 2005. **333**(4): p. 1309-14.

134. Hosking, L., et al., *Detection of genotyping errors by Hardy-Weinberg equilibrium testing.* Eur J Hum Genet, 2004. **12**(5): p. 395-9.

135. Wittke-Thompson, J.K., A. Pluzhnikov, and N.J. Cox, *Rational inferences about departures from Hardy-Weinberg equilibrium.* Am J Hum Genet, 2005. **76**(6): p. 967-86.

136. Fredman, D., et al., *Complex SNP-related sequence variation in segmental genome duplications.* Nat Genet, 2004. **36**(8): p. 861-6.

137. Lin, M., et al., *dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data.* Bioinformatics, 2004. **20**(8): p. 1233-40.

138. Laframboise, T., D. Harrington, and B.A. Weir, *PLASQ: a generalized linear model-based procedure to determine allelic dosage in cancer cells from SNP array data.* Biostatistics, 2007. **8**(2): p. 323-36.

139. Altschul, S.F., et al., *Basic Local Alignment Search Tool.* J. Mol. Biol., 1990. **215**: p. 403-410.

140. Zhang, J., et al., *Development of bioinformatics resources for display and analysis of copy number and other structural variants in the human genome.* Cytogenet Genome Res, 2006. **115**(3-4): p. 205-14.

141. Eddy, S.R., *What is a hidden Markov model?* Nat Biotechnol, 2004. **22**(10): p. 1315-6.

142. Olshen, A.B., et al., *Circular binary segmentation for the analysis of array-based DNA copy number data.* Biostatistics, 2004. **5**(4): p. 557-72.

143. Tuzun, E., et al., *Fine-scale structural variation of the human genome.* Nat Genet, 2005. **37**(7): p. 727-32.

144. Sebat, J., et al., *Strong association of de novo copy number mutations with autism.* Science, 2007. **316**(5823): p. 445-9.

145. Friedman, J.M., et al., *Oligonucleotide microarray analysis of genomic imbalance in children with mental retardation.* Am J Hum Genet, 2006. **79**(3): p. 500-13.

146. Abecasis, G., et al., *Human Genome Variation 2006: emerging views on structural variation and large-scale SNP analysis.* Nat Genet, 2007. **39**(2): p. 153-5.

147. Locke, D.P., et al., *Linkage Disequilibrium and Heritability of Copy-Number Polymorphisms within Duplicated Regions of the Human Genome.* Am J Hum Genet, 2006. **79**(2): p. 275-90.

148. Stern, S., et al., *Genome-wide transcriptional plasticity underlies cellular adaptation to novel challenge.* Mol Syst Biol, 2007. **3**: p. 106.

149. Kraft, P., *Multiple comparisons in studies of gene x gene and gene x environment interaction.* Am J Hum Genet, 2004. **74**(3): p. 582-4; author reply 584-5.

Appendix:

Integration of Methods: Application of Multiple Attribute Decision Making

I. Description of the research question:

Neurodegeneration due to Alzheimer's disease (AD) is the most frequent cause of cognitive decline in the elderly and afflicts approximately 26 million individuals worldwide[59]. The common, late-onset form of AD is a multifactorial disease involving multiple cellular processes and environmental influences. Linkage and whole genome association studies show evidence for genes on 11 chromosomes[60]. Choosing specific genes from among the many possible candidates within the linkage region for followup association studies relies heavily on prior biological knowledge and assumptions. However, these methods have had limited success in clearly identifying susceptibility genes. Using gene expression results to inform decisions of candidate gene association studies for AD would improve our ability to choose rational candidates. We used 9 clinically non-demented and 5 AD cases to investigate differential gene expression across the entire human genome using Affymetrix HG-U133Plus2 GeneChips (manuscript submitted) and determine genes involved in AD pathology. In order to prioritize genes for subsequent candidate gene studies in an unbiased fashion, we devised a Multiple Attribute Decision Making (MADM) strategy to create a disease score for integrating empirical data from diverse biological methods.

II. MADM method in AD

The goal is to prioritize all of the genes on the GeneChip in terms of their relative probability of success in future candidate gene association studies. We wanted to select among differentially expressed genes in a formalized framework for judging important biological information.

Alternatives: we defined all of the genes on the GeneChip as the set of alternatives from which we wanted to chose a final list of genes for further candidate gene studies.

Attributes: the single most important feature of a MADM analysis is the choice of attributes. The types of information pertaining to the problem of AD pathology were organized in a hierarchical fashion. At the top is the highest goal: to prioritize genes for further studies. Secondly, we wanted to capture as many novel genes as possible. Subattributes were chosen to be independent from one another. Choices that did not provide new knowledge were discarded due to overlapping information content. We chose gene expression q-value, presence in a known linkage/ association region for AD, and overexpression in the Gene Ontology (GO) Biological Process category (http://www.geneontology.org).

Units of Measure: Attributes in which the higher the value corresponds to a greater preference were counted as 1or 0 such as 1) presence of a gene in a linkage region and 2) overrepresentation in a GO category. Each gene with a measurement of 1 for these attributes was counted in the final MADM score. The q-value is quantitative and more significant at lower values than the higher values. The lower the score is, the greater the preference. To allow for combining with other attributes so that the higher values correspond to greater preference, q-value was normalized by:

$$\min(j) / x_{ij}$$

In this case, min(j) is the minimum of all q-values and x_{ij} is the q-value(j) for every gene i.

Attribute Weights: Each attribute was weighted based on the importance of the biological information assessed. We considered presence in a linkage /association region to be the most important criterion for inclusion in a candidate gene study and was given a weight of 0.5. Overrepresentation in the GO biological Process category was weighted as 0.3 and q-value was used to prioritize the biological information by significance of gene expression (0.2). The final decision hierarchy is shown in Figure 1. The final score for each gene was the sum of the values across the attributes for each genes multiplied by their weights.

$$\mathrm{Vi} = \sum_{j=1}^{n} w_j x_{ij}$$

where w is the weight for each attribute j and x_{ij} is the value (or normalized value) for each gene (i) at that attribute j.

III. Results

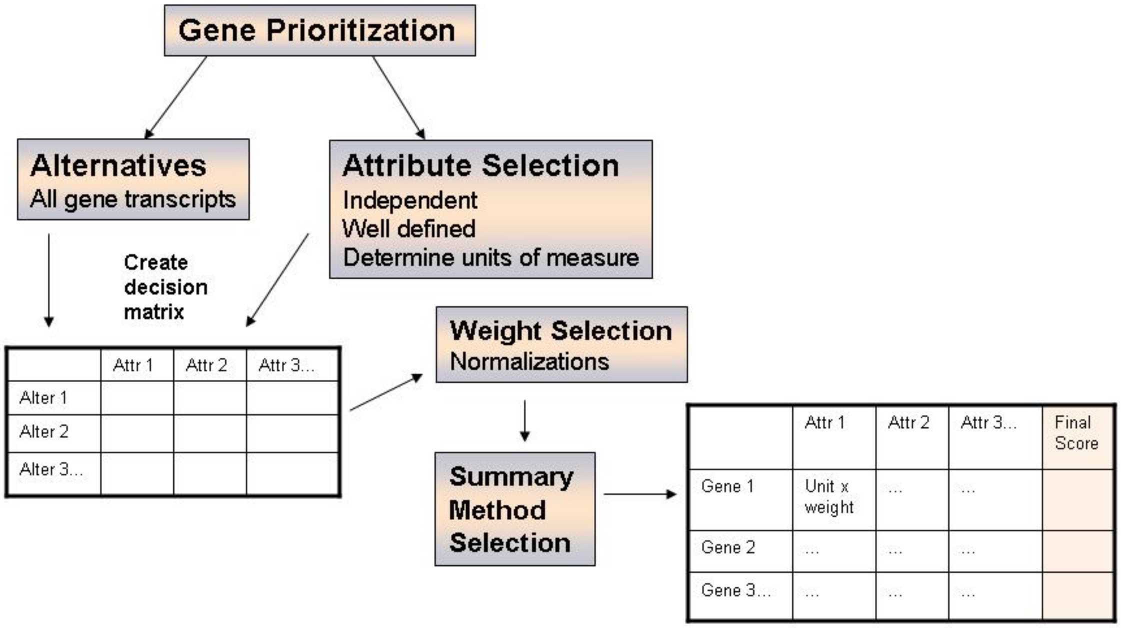
For this weighting scheme, the closer the MADM score is to 1 the better it is. The final gene list included 210 genes with a score > 0.8. The highest score was 0.922. Genes on this list all had a q-value < 0.50 and 35 genes had a q-value < 0.01. All genes were located in AD linkage/association regions and were overrepresented in GO Biological Process categories. The difference in the MADM gene list and the original gene list lies in the

IV. Discussion

The difference in the MADM gene list and the original gene list lies in the conscious decision to require genes of interest to be located in AD linkage/association regions and overrepresentation in GO Biological Process categories. The original gene list was sorted by q-value and interesting genes identified by manual curation. MADM methods allow for subsequent research objectives to be precisely defined. Candidate gene studies based on the MADM list would reflect the predefined criteria as determined by the problem of interest.

The choice of problem description should be general enough to capture the appropriate choice and specific enough to define independent attributes. In our study, we were most interested in discovery of novel genes involved in AD pathology. Other possible questions would necessitate additional attributes and weighting schemes or new hierarchies to be created. For example, to identify a combination of novel genes and previously studied AD genes, additional attributes would be added to the hierarchy and weights would be modified to reflect the importance of gene discovery versus gene confirmation. A focus on healthy cognitive aging within the comparisons of the experiment, would necessitate the creation of different hierarchies.

Figure 1. MADM decision heirarchy.



	Attr 1	Attr 2	Attr 3…
Alter 1			
Alter 2			
Alter 3…			

	Attr 1	Attr 2	Attr 3…	Final Score
Gene 1	Unit x weight	…	…	
Gene 2	…	…	…	
Gene 3…	…	…	…	