EUROPEAN JOURNAL OF INFLAMMATION Vol. 12, no. 2, 267-276 (2014)

GENOME-WIDE SCAN FOR COPY NUMBER VARIATION ASSOCIATION WITH BIOMARKER QUANTITATIVE TRAIT LOCI IN AGING

K. SZIGETI[1], B. TRUMMER[1], D. LAL[1], R.S. DOODY[2], L. YAN[3], S. LIU[4] and C. MA[3]
for the Texas Alzheimer Research and Care Consortium and the Alzheimer's Disease Neuroimaging Initiative*

[1]Department of Neurology, University at Buffalo, SUNY, Buffalo, NY, USA; [2]Alzheimer's Disease and Memory Disorders Center, Department of Neurology, Baylor College of Medicine, Houston, TX, USA; [3]Department of Bioinformatics, University at Buffalo, SUNY, Buffalo, NY, USA; [4]Roswell Park Cancer Institute, Buffalo, NY, USA

Received June 4, 2013 – Accepted March 25, 2014

Biomarkers are emerging as important tools in the detection and monitoring of various diseases. A major limitation and challenge to effectively utilize biomarker signals is the limited understanding of factors contributing to their variance. As genetic variation is a major contributor to phenotypic variation, exploring genetic contributions is of great importance. Copy number variants (CNVs) offer an alternative genomic framework to understand contributions to phenotypic variance. A copy-number variation genome-wide association study was performed using 116 serum inflammatory biomarkers as quantitative trait in elderly normal controls to test the hypothesis that CNVs contribute to the phenotypic heterogeneity of serum biomarkers. Three chromosomal regions were associated with four biomarkers in trans. Transforming growth factor alpha (TG-alpha) serum levels were associated with CNV dosage at chr11:5,788 kb, soluble levels of receptor for advanced glycation endproducts (sRAGE) was associated with CNV dosage at chr8:40,183 kb and both thrombospondin-1 and tissue inhibitor of metalloproteinase 1 (TIMP-1) were associated with CNV dosage at chr11:18,961 kb. The CNV at chr11:5,788 kb harbors 2 olfactory genes and the introns of Tripartite motif-containing (TRIM) gene cluster TRIM5&22 while the CNV at chr11:18,961 includes the Mas-related G-protein coupled receptor member X1. These trans associations may identify novel relationships in the relevant pathways and suggest that genetic variation can contribute to biomarker levels. The detected trans-association between *MRGPRX1* and thrombospondin-1/TIMP-1 could implicate a novel pathway between pain/itching and inflammation. Cataloguing all genetic variants with an effect on biomarkers will serve as a tool to interpret epidemiological studies and establish causal relationships through Mendelian randomization.

The observation of widespread alteration in the copy number of submicroscopic DNA segments has contributed to the complexity of human genetic variation (1). Copy number variants (CNV) influence gene expression, phenotypic variation and adaptation by altering gene dosage and/or genome organization (2). CNVs are a group of structural variants and can be classified as deletions, duplications, deletions and duplications at the same locus, and complex rearrangements (3). CNVs are often multiallelic

Key words: copy number variation, biomarker, aging

Mailing address: Dr Kinga Szigeti,
University of Buffalo SUNY,
100 High Street,
Buffalo, NY 14203, USA
Tel.: +1 716 859 3484
Fax: +1 716 859 7833
e-mail: szigeti@buffalo.edu

1721-727X (2014)

DISCLOSURE: ALL AUTHORS REPORT NO CONFLICTS OF INTEREST RELEVANT TO THIS ARTICLE.

and have a higher new mutation rate than single nucleotide polymorphisms (SNP) (4). Due to these properties, CNVs offer a complementary genetic marker map to SNPs.

Biomarkers are measurable indicators of biological processes, and are used to diagnose disease or assess treatment response (5). Biomarkers in addition may suggest specific pathways for intervention (5). Three major challenges of proteomic biomarker discovery include the complexity of the various proteins in the plasma, the low abundance necessitating sensitive methods to detect specific biomarkers, and the extent of human variation and subsequent disease variation (5). Understanding the association between genetic variation and biomarker levels may refine associations and establish causal relationships in epidemiological studies through Mendelian randomization (6-8). Traditional biomarker-disease association studies cannot establish causation and are frequently confounded (7, 8). If a genetic marker is identified that has an effect on the intermediate factor (biomarker), the observational studies can be further interpreted through Mendelian randomization; by using the risk allele the study is randomized at conception and the direction of effect (variant altering intermediate factor) is known (7, 8). In order to use a given genetic variation for this application, multiple independent studies need to confirm the association between risk allele and biomarker levels (7).

CNV association with expression quantitative trait loci has been studied in cell lines and approximately 18% of gene expression traits are associated with CNVs (2). CNVs are major contributors to genetic variance thus it is conceivable that they may contribute to the heritability of biomarker levels (9-11). To test this hypothesis, we performed a genome-wide association study using CNVs as a genetic marker map and a panel of inflammatory biomarkers as a quantitative trait in a normal aging cohort. To our knowledge this is the first study to explore CNV-biomarker associations at the whole genome level.

MATERIALS AND METHODS

Ethics statement

The normal aging study of the Texas Alzheimer's Disease Research and Care Consortium (TARCC) enrolled 240 control subjects after written informed consent. The IRB of each participating site, Texas Tech University Health Science Center, University of North Texas Health Science Center, the University of Texas Southwestern Medical Center at Dallas, and Baylor College of Medicine approved the study.

Subject cohorts

The methodology of the TARCC project has been described in detail elsewhere (12). Inclusion criteria briefly included age over 55 years, male and female, and normal cognitive function and activities of daily living. All control subjects underwent neuropsychological testing including assessment of Global cognitive functioning/status (MMSE and CDR), Attention (Digit Span and Trails A), Executive function (Trails B and Clock Drawing; Texas Card Sorting is optional), Memory (WMS Logical Memory I and WMS Logical Memory II), Language (Boston Naming and FAS Verbal Fluency), Premorbid IQ (AMNART), Visuospatial Memory (WMS-Visual Reproduction I and II), Psychiatric symptoms (Geriatric Depression Scale; Neuropsychiatric Inventory-Questionnaire) and Functional assessment (Lawton-Brody ADL: PSMS, IADL). All scores had to be within normal range to be included in the study.

Serum biomarkers

Non-fasting blood samples were collected in serum-separating tubes, allowed to clot at room temperature for 30 min, centrifuged, aliquoted, and stored at -80°C in plastic vials. Batched specimens were sent frozen to Rules Based Medicine (RBM, www.rulesbasedmedicine.com, Austin, TX) for protein quantification by multiplex fluorescent immunoassay (Multi-Analyte Profile (humanMAP)) utilizing colored microspheres with protein-specific antibodies. Quality control parameters including the least detectable dose (LDD), inter-run coefficient of variation, dynamic range, overall spiked standard recovery and cross-reactivity with other human MAP analytes are available at Rules Based Medicine.

CNV genotyping by the Genome-wide human SNP array 6.0

Genomic DNA was isolated from whole blood by the Puregene DNA isolation kit (Qiagen) according to the manufacturer's instructions. Array based genotyping was performed on the Genome-Wide Human SNP Array 6.0 (Affymetrix) according to the manufacturer's instructions. The following QC measures were applied: contrast QC (>0.4) and Median of the Absolute values of all Pairwise Differences (MAPD) < 0.4.

Detection of copy number variation and test of association

Principal component analysis and correction was performed using the Eigenstrat method (13) implemented in Goldenhelix (GoldenHelix). The PCA corrected data

was segmented by the CNAM algorithm implemented in GoldenHelix (GoldenHelix). The algorithm scans the probes aligned to the genome where a significant break occurs and identifies it as a start or end of a copy number event. The univariate method identifies rare CNVs by scanning the data subject by subject. In the univariate method a moving window of 20000 probes was applied. Maximum and minimum number of segments per 10000 markers was set at 10 and 1, respectively. The maximum pairwise permuted p-value was determined at 0.005, and 2000 permutations per pair were applied. The multivariate method compares all samples at each locus and identifies common CNVs. In the multivariate segmentation we did not use a moving window; otherwise the parameters were kept the same. The goal is to have high sensitivity to detect association signals from small events with the expectations that all associations will be manually reviewed for validation. The segmentation covariates generated by either algorithm was used in the test of association as predictor.

For test of association we grouped the biomarkers into three categories based on the distribution of each biomarker dataset as this approach prevented the inflation of the test statistics. The first set has normal distribution and we applied a linear regression model using the segmented CNV data and the biomarker data in the regression. The second group of biomarkers was normally distributed after log transformation. In this set we used the log transformed biomarker data and the segmented CNV data in the linear regression model. The third set of biomarkers had non-normal distribution; for this set we used a nonparametric test of association (Spearman rank test) and we used the segmented CNV data and the raw biomarker data in the model. The FDR approach was used for multiple testing correction for both the multivariate and univariate segments in relationship with all 116 biomarkers. The FDR was calculated based on all p-values from every biomarker association. We did not correct for the 2 segmentation algorithm as there is marked overlap between the covariates and thus it would have been overly conservative. The linear regression model was repeated for the significant associations with age and sex added to the model.

Replication of the results

The only existing dataset with comparable inflammatory biomarker and CNV data is the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. Normal control biomarker and genome-wide genetic data were obtained from the ADNI database (adni.loni.ucla.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year public private partnership. The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California, San Francisco, USA. ADNI is the result of efforts of many coinvestigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. For up-to-date information, see www.adni-info.org.

The ADNI genotyping was performed on the Illumina Human610-Quad Bead Chip. The logR ratio was calculated in GenomeStudio (Illumina) software. Candidate locus specific data was extracted and the CNV calls were based on the logR distributions by inspection of the kernel density estimate of logR and identification of cutpoints that distinguish the CNV dosages (loss, normal, gain). For each CNV biomarker pair the same test of association was performed as in the discovery cohort. The biomarker dataset was plasma based in ADNI compared to serum levels in the discovery cohort. The TARCC performed a comparison between the serum and plasma biomarker levels from the same individuals at the same time; only those biomarkers that had a serum-plasma correlation $r^2 \geq 0.8$ were considered fit for replication by the ADNI set.

Validation of CNV calls

Real-time PCR (Taqman assay) was performed according to the manufacturer's instructions (Applied Biosystems, Inc). The duplex real-time PCR assays were performed using a Fam dye labeled assay targeted to Chromosome 11:18,961 kb (Assay ID: Hs01727613_CN), Chromosome 8:40,183 kb (Assay ID:Hs03268981_CN) and Chromosome 11:5,788 (Assay ID:Hs03774711_CN and the calibrator RNase P-primers with a VIC-dye labelled probe (TaqMan copy number reference assay, part # 4403326) as a reference gene. Each sample was assayed in quadruplicate by using 10 ng DNA in each reaction. Real-time PCR was performed using the CFX384 Real-time PCR Detection System (Bio-Rad). Threshold cycle (Ct) values were determined for CHRFAM7A and compared with Ct values for RNase P. Relative quantity was determined by the DD Ct method (14).

RESULTS

Two hundred forty samples passed the intensity filter default (0.4) setting in the genotyping Concole (Affymetrix) and were adequate for copy number analysis. Thirteen samples failed MAPD cutoff of 0.4 and 35 samples failed because the intensity data

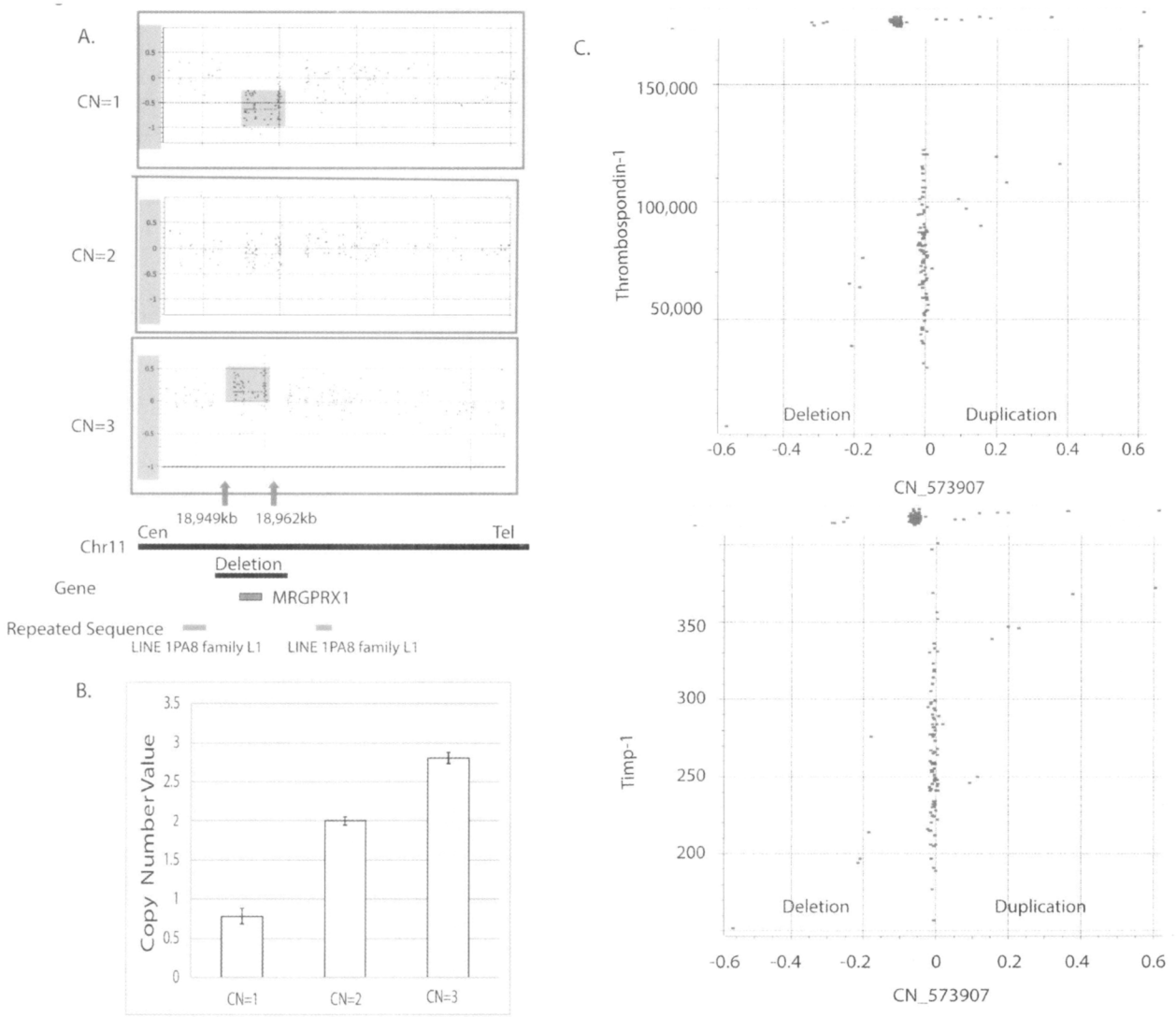

Fig. 1. *Sample array data for the Chr 11:5,788 kb CNV events. Intragenic deletion of TRIM5/TRIM22 introns and the two olfactory genes OR52N5 and OR52N1 is detected. The samples demonstrate homozygous (0 copy) and heterozygous (1 copy) deletions (**A**). Validation of the 3 CN (CN=0, 1 and 2) states using TaqMan Copy Number Assay (**B**). The scatter plot represents the TGF-alpha (y axis) and CN_568120 segmentation covariate association (**C**). The CNV segmentation covariates depict an elevated level of TGF-alpha if a deletion is present and marked increase in TGF-alpha levels if the CNV is in the homozygous deletion state.*

distribution exhibited an excess number of CNV calls more than 2SD of the mean. Overall, 192 subject samples passed QC. Out of the 192 controls 106 had biomarker data. Demographics of the 106 subjects are depicted in Table I.

One hundred sixteen biomarkers passed QC criteria as they were detected above LDD. Biomarker data normality testing resulted in 10 biomarkers that were normally distributed, 45 biomarkers that were normally distributed after log transformation and 61 biomarkers that were non-normally distributed.

In the first set (normal distribution, linear regression) the linear regression model identified a univariate segment at chr11:18,961 kb where the association signal with thrombospondin-1 survived multiple testing correction and the association signal with tissue inhibitor of metalloproteinase 1 (TIMP-1) showed a trend with the same CNV within the same

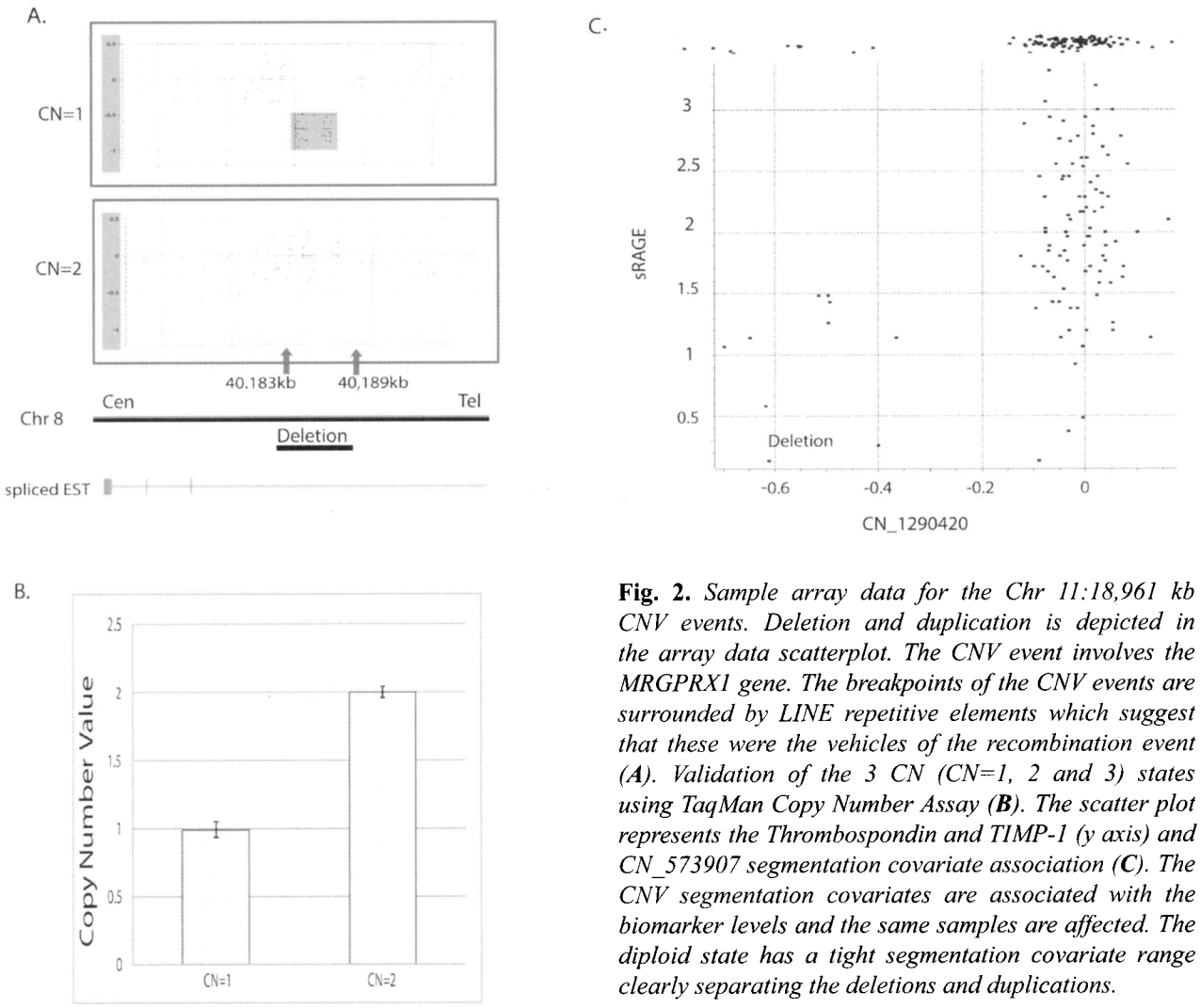

Fig. 2. *Sample array data for the Chr 11:18,961 kb CNV events. Deletion and duplication is depicted in the array data scatterplot. The CNV event involves the MRGPRX1 gene. The breakpoints of the CNV events are surrounded by LINE repetitive elements which suggest that these were the vehicles of the recombination event (**A**). Validation of the 3 CN (CN=1, 2 and 3) states using TaqMan Copy Number Assay (**B**). The scatter plot represents the Thrombospondin and TIMP-1 (y axis) and CN_573907 segmentation covariate association (**C**). The CNV segmentation covariates are associated with the biomarker levels and the same samples are affected. The diploid state has a tight segmentation covariate range clearly separating the deletions and duplications.*

samples (Table II). The biomarker and segmented logR scatterplots validated the association signal (Fig. 1). In the second set (normal distribution of log transformed biomarker data, linear regression) the linear regression model identified two multivariate segments on chr8:40,183 kb and chr11:5,788 kb (Table II). The CNV at chr8:40,183 kb was associated with sRAGE levels and the CNV at chr11:5,788 kb with TGF-alpha serum levels. Scatterplots validated the association signal (Figs. 2 and 3). In the third set (non-normal distribution, nonparametric test) the Spearman rank test identified three multivariate segments. Manual review of the CNV events and the biomarker versus segmented logR scatterplots showed that two of these associations are spurious and not related to true CNV calls. The third association serves as an internal positive control: the association signal with FSH originates from a segment located on chr1:241,116 kb. This CNV is in fact a low-copy repeat with a 99% identical copy on chromosome Y. The association between FSH levels and dosage correlates with gender (Fig. 4).

The number of events contributing to the association signals are depicted in Table II and reflect an allele frequency of the CNVs between 9-25%. The allele frequencies are consistent with the sample size and it is unlikely that these are spurious associations. All regions are known CNV regions. The detected CNVs were small, ranging from 5.7-21 kb. The four trans associations remained significant after

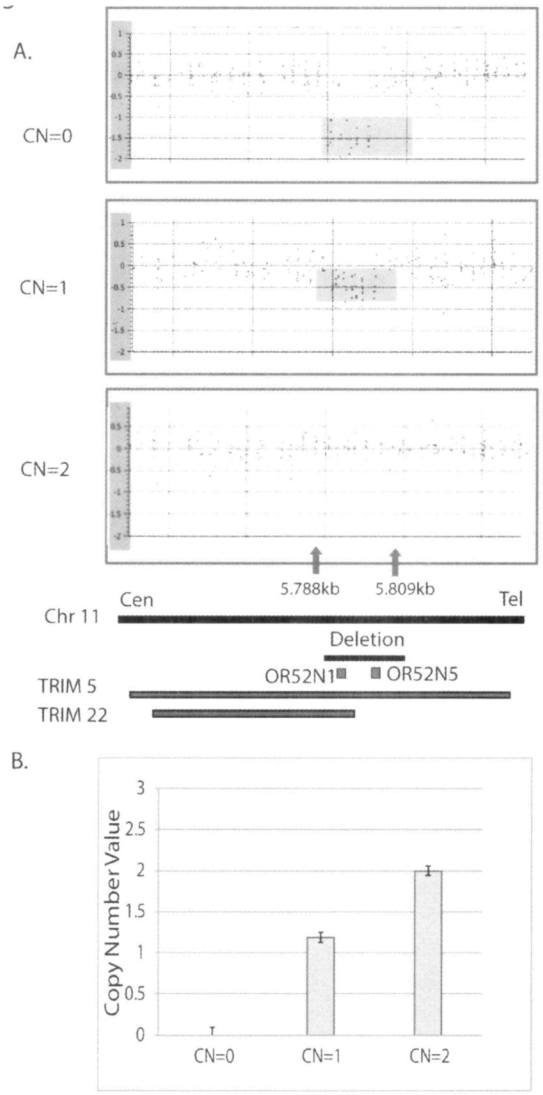

Fig. 3. *Sample array data for the chr8:40,183 kb CNV events. The deletion harbors spliced ESTs and lamin B1 binding sites (**A**). Validation of the 2 CN (CN=1and 2) states using TaqMan Copy Number Assay (**B**). The scatter plot represents the soluble RAGE (y axis) and CN_1290420 segmentation covariate association (**C**). The CNV segmentation covariates depict a decreased level of sRAGE if a deletion is present.*

adding age and sex to the linear regression models (Table II). TIMP-1 and sRAGE had a significant correlation with age, as expected (p=0.03 and 0.009, respectively), but the CNV association remained significant after correcting for this variable. CNV events detected on the Affymetrix platform were validated by TaqMan assay (Figs. 1B, 2B and 3B).

The ADNI study was evaluated for comparability with the goal to use it as a replication study. As the ADNI biomarkers were measured in plasma and the TARCC biomarkers in serum, first we assessed the correlation of biomarker levels between serum and plasma in 40 TARCC subjects. The serum-plasma correlation for the candidate biomarkers showed an r^2 0.1 for TIMP-1, 0.038 for thrombospondin and 0.8 for sRAGE. TGF-alpha was undetectable in most of the ADNI plasma samples.

There were 63 ADNI controls where biomarker and Illumina data was available, thus the replication study was underpowered. The 63 ADNI controls were studied for replication of the sRAGE chr8:40,183 kb association. Seven probes were available in the chr8:40,183 kb region on the Illumina 610 array, two of which had a kernel distribution that allowed CNV calls. The association with the two best performing probes showed a trend: the uncorrected p-value for rs210658 and cnvi0001732 association with sRAGE was 0.06 and 0.13, respectively.

DISCUSSION

The genomic screen for CNV and biomarker as

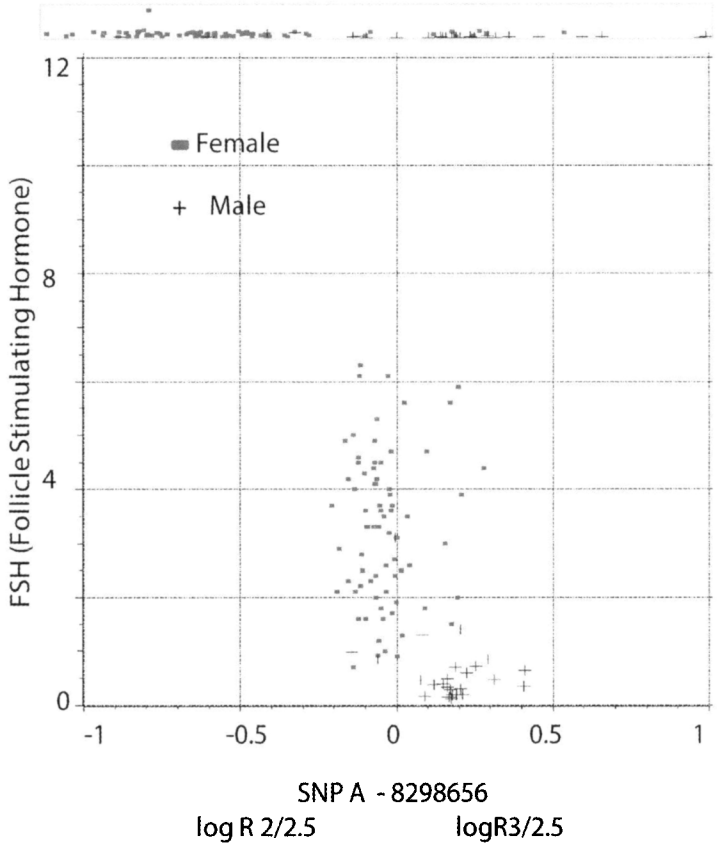

Fig. 4. *The scatter plot represents the FSH (y axis) and SNP_8298656 segmentation covariate association split by_sex. SNP_8298656 is located on chr1:243,049kb, in an LCR that has the homologues sequence on chromosome Y. As the reference set is calculated as approximately same number of males and females, the denominator for the logR calculation is 2.5. In males there are 3 copies, 1 on chromosome Y and two on chromosome 1, while in females there is two copies on chromosome 1. The logR is calculated as log2(3/2.5) for males and log2(2/2.5) for females, which corresponds to 0.26 and -0.32 on the Y axis.*

Table I. *Demographics of the discovery cohort.*

Number of subjects	106
Female	79 (74%)
Age	73 (+/-8.95)
Hyperlipidemia	48 (45%)
Hypertension	64 (60%)
Obesity	31 (29%)
Diabetes	9 (0.8%)
Mini Mental Status Exam	29(+/-0.87)
sRAGE	4.29(+/-1.9)
TGF-alpha	66.14(+/-36.5)
Thrombospondin	79326(+/-24535)
TIMP-1	215.65(+/-50.1)

a quantitative trait association study detected four trans associations, chr11:18,961 kb association with thrombospondin-1 and TIMP-1, chr8:40,183 kb association with sRAGE and chr11:5,788 kb association with TGF-alpha. The associations remained significant after adding age and sex to the linear regression models.

The detected events ranged from 5.7-21 kb. Most CNV association studies to date applied segmentation followed by high stringency calls often over 100kb or over 10 probes contributing to the CNV events. While this approach has high specificity for large events, it has very low sensitivity for smaller events. The strategy applied here, using segmentation covariates without binning, made the statistics more robust, and manual review eliminated the false

Table II. *Trans CNV and biomarker associations detected in the discovery cohort.*

Set	Segmentation	Biomarker	CNV/SNP	Chr	Location	p-value	FDR pvalue	p-value with age and sex	Number events	Frequency	Number probes	Size	Gene
1	Univariate	Thrombospondin-1	CN_573907	11	18961976	3.98×10^{-9}	0.0003	3.96×10^{-9}	12	0.11	56	13.3kb	MRGPRX1
1	Univariate	TIMP-1	CN_573907	11	18961976	7.61×10^{-7}	0.0599	1.96×10^{-6}	12	0.11	56	13.3kb	MRGPRX1
2	Multivariate	sRAGE	CN_1290420	8	40183785	3.65×10^{-6}	0.0066	2.08×10^{-5}	10	0.09	31	5.7kb	none
2	Multivariate	TGF-alpha	CN_568120	11	5788068	2.94×10^{-6}	0.0053	3.02×10^{-6}	26	0.25	31	21 kb	TRIM5, TRIM22, OR52N5, OR52N1

positives. The detected events are all known CNVs, are consistently detected between samples with 31-56 probes and were validated by an orthogonal method.

During the efforts to replicate the association results we identified the ADNI dataset as a potential replication cohort. The cohort focuses on the geriatric population, has normal controls, GWAS data is available on the Illumina Human610-Quad Bead Chip and RBM biomarker data is available. However, there are several limitations to use this cohort as a replication cohort: the biomarkers were measured in plasma samples while the TARCC set was in serum. The sample size of the normal Caucasian controls with both biomarker and genome-wide association study (GWAS) data in the ADNI study was 63, which makes the replication study underpowered. In order to overcome the serum versus plasma problem, we ran 40 TARCC subject samples concomitantly as plasma and as serum. Only sRAGE demonstrated a correlation between serum and plasma with an r^2 of 0.8. Due to these limitations the only association for potential replication was the sRAGE and chr8:40,183 kb association. Despite the markedly underpowered replication cohort, the uncorrected p-value for the association of the best performing Illumina probes rs210658 and cnvi0001732 with sRAGE showed a trend. Further studies are needed to replicate this and the other trans associations. To our knowledge, there are no other studies where RBM biomarker and GWAS data on the geriatric population are available.

Two of the detected CNVs are intragenic: the CNV at chr11:18,961 kb harbors Mas- related G-protein coupled receptor member X1 (*MRGPRX1*), which is expressed in specific subsets of nociceptive sensory neurons and is implicated in the sensation of itching (15, 16). Thrombospondin-1 is an adhesive glucoprotein that mediates cell-cell and cell-matrix interactions (17). Thrombospondin-1 can bind to some of the matrix metalloproteinases (MMPs) and assist in their removal from the extracellular environment (17). TIMP-1 is a tissue inhibitor of the MMPs and its expression is regulated by various cytokines and hormones (18). The detected trans-association between *MRGPRX1*and thrombospondin-1/TIMP-1 could implicate a novel pathway between pain/itching and inflammation. Further molecular work and translational studies are needed to fully evaluate this relationship.

The CNV at chr8:40,183 kb does not have any known genes within the CNV event; however, there are spliced ESTs (e.g. HY042748) and several conserved segments. This CNV also has multiple LaminB1 binding sites. The relationship of this chromosomal region to sRAGE levels is unclear at this point.

The CNV at chr11:5,788 kb harbors the introns of *TRIM5* and *TRIM22*, and two complete olfactory receptor genes, Olfactory receptor, family 52, subfamily N, members 5 and 1 (*OR52N5)* and (*OR52N1)*. This chromosomal region is an interesting CNV structure: a TRIM cluster is embedded within the olfactory supercluster. The relationship between TGF-alpha levels and the gene dosage of the TRIM cluster or the OR cluster is plausible, although not yet known. TRIM proteins are involved in innate immunity, retroviral restriction, antiviral defense and autoimmunity. TRIM5 and 22 belong to the C-IV family and have E3 ubiquitin ligase activity (19). TRIM5 targets both TGF-beta activated kinase 1/MAP3K7 binding protein 2 and 3 (TAB2) and (TAB3) for degradation, thus abrogates the activation

of TGF-beta activated kinase 1 and the subsequent downstream activation of NF-κB (19). Another TRIM protein, TRIM33, is a chromatin reader and aids the access of SMAD to chromatin in the TGF-beta canonical pathway (20, 21). Further work is needed to evaluate whether there is a relationship between TGF-alpha and allelic variants of TRIM. The olfactory receptor CNV association with TGF-alpha has not been reported yet; however TGF-alpha has a role in olfactory recovery after nerve transection and stimulates cell division in the olfactory epithelium.

These CNV and biomarker associations may suggest novel interactions between pathways that were previously unlinked. Further larger scale studies are needed to fully investigate these trans-relationships in normal and disease groups and to elucidate the precise biochemical pathways underlying the association. Cataloguing all genetic variants with an effect on biomarkers will serve as a tool to interpret epidemiological studies and establish causal relationships through Mendelian randomization.

ACKNOWLEDGEMENTS

This study was supported by the Texas Alzheimer's Research and Care Consortium (TARCC) and an Alzheimer Association New Investigator Research Grant to KS.

*Investigators from the Texas Alzheimer's Research and Care Consortium: Baylor College of Medicine: S. Rountree, V. Pavlik, W, Chan, P. Massman, E. Darby, T. Evans RN, A. Khaleeq; Texas Tech University Health Science Center: B. Williams, G. Schrimsher, A. Dentino, R. Orozco; University of North Texas Health Science Center: T. Fairchild, J. Knebl, S.E. O'Bryant, James R. Hall, Robert C. Barber, D. Mains, L. Alvarez; University of Texas Southwestern Medical Center: P. Adams, R. Rosenberg, M. Weiner, M. Quiceno, J. Reisch, R.Huebinger, G. Xiao, D. Svetlik, A. Werry, J. Smith; University of Texas Health Science Center – San Antonio: D. Royall, R. Palmer, M.Polk.

Data collection and sharing for part of this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904). Data used in preparation of this article were obtained from the ADNI database (adni.loni.ucla.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

REFERENCES

1. Redon R, Ishikawa S, Fitch KR, et al. Global variation in copy number in the human genome. Nature 2006; 444:444-54.
2. Stranger BE, Forrest MS, Dunning M, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science 2007; 315:848-53.
3. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. Nat Rev Genet 2006; 7:85-97.
4. Conrad DF, Bird C, Blackburne B, Lindsay S, Mamanova L, Lee C, Turner DJ, Hurles ME. Mutation spectrum revealed by breakpoint sequencing of human germline CNVs. Nat Genet 2010; 42:385-91.
5. Rifai N, Gillette MA, Carr SA. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. Nat Biotechnol 2006; 24:971-83.
6. Berry DJ, Vimaleswaran KS, Whittaker JC, Hingorani AD, Hypponen E. Evaluation of genetic markers as instruments for Mendelian randomization studies on vitamin D. PLoS One 2012; 7:e37465.
7. Palmer TM, Thompson JR, Tobin MD, Sheehan NA, Burton PR. Adjusting for bias and unmeasured confounding in Mendelian randomization studies with binary responses. Int J Epidemiol 2008; 37:1161-8.
8. Vuistiner P, Bochud M, Rousson V. A comparison of three methods of Mendelian randomization when the genetic instrument, the risk factor and the outcome are all binary. PLoS One 2012; 7:e35951.
9. Guthrie PA, Rodriguez S, Gaunt TR, Lawlor DA, Smith GD, Day IN. Complexity of a complex trait locus: HP, HPR, haemoglobin and cholesterol. Gene 2012; 499:8-13.
10. Kim HY, Cho S, Yu J, Sung S, Kim H. Analysis of copy number variation in 8,842 Korean individuals reveals 39 genes associated with hepatic biomarkers

AST and ALT. BMB Rep 2010; 43:547-53.
11. Kim HY, Byun MJ, Kim H. A replication study of genome-wide CNV association for hepatic biomarkers identifies nine genes associated with liver function. BMB Rep 2011; 44:578-83.
12. Waring SC OBS, Reisch J, Diaz-Arrastia R, Knebl J, Doody RS. The Texas Alzheimer Research Consortium longitudinal research cohort: study design and baseline characteristics. J Texas Public Health Assoc 2008; 60:10-13.
13. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NAReich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006; 38:904-9.
14. Aldhous MC, Abu Bakar S, Prescott NJ, et al. Measurement methods and accuracy in copy number variation: failure to replicate associations of beta-defensin copy number with Crohn's disease. Hum Mol Genet 2010; 19:4930-38.
15. Liu Q, Tang Z, Surdenikova L, et al. Sensory neuron-specific GPCR Mrgprs are itch receptors mediating chloroquine-induced pruritus. Cell 2009; 139:1353-65.
16. Sikand P, Dong X, LaMotte RH. BAM8-22 peptide produces itch and nociceptive sensations in humans independent of histamine release. J Neurosci 2011; 31:7563-7.
17. Lopez-Dee Z, Pidcock KGutierrez LS. Thrombospondin-1: multiple paths to inflammation. Mediators Inflamm 2011; 2011:296069.
18. Moore CS, Crocker SJ. An alternate perspective on the roles of TIMPs and MMPs in pathology. Am J Pathol 2012; 180:12-6.
19. Ozato K, Shin DM, Chang TH, Morse HC, 3rd. TRIM family proteins and their emerging roles in innate immunity. Nat Rev Immunol 2008; 8:849-60.
20. Massague J. TGF-beta signaling in context. Nat Rev Mol Cell Biol 2012; 13:616-30.
21. Massague J. TGF-beta signaling in development and disease. FEBS Lett 2012; 586:1833.