PI: **Szigeti, Kinga**	Title: Narrowing the gap in the genetic architecture of Alzheimer's disease	
Received: 11/05/2013	FOA: PA13-302	Council: 05/2014
Competition ID: FORMS-C	FOA Title: RESEARCH PROJECT GRANT (PARENT R01)	
1 R01 AG046345-01A1	Dual:	Accession Number: 3636703
IPF: 5992614	Organization: STATE UNIVERSITY OF NEW YORK AT BUFFALO	
Former Number:	Department: Neurology	
IRG/SRG: GHD	AIDS: N	Expedited: N
Subtotal Direct Costs (excludes consortium F&A) Year 1: 250,000 Year 2: 250,000 Year 3: 250,000 Year 4: 250,000 Year 5: 250,000	Animals: N Humans: Y Clinical Trial: N Current HS Code: E4 HESC: N	New Investigator: Y Early Stage Investigator: Y
Senior/Key Personnel:	*Organization:*	*Role Category:*
Kinga Szigeti	State University of New York at Buffalo	PD/PI
Li Yan Ph.D	Roswell Park Cancer Institute	Other (Specify)-Senior Biostatistician
Jianmin Wang Ph.D	Roswell Park Cancer Institute	Co-Investigator

OMB Number: 4040-0001
Expiration Date: 06/30/2016

APPLICATION FOR FEDERAL ASSISTANCE
SF 424 (R&R)

3. DATE RECEIVED BY STATE

State Application Identifier

1. TYPE OF SUBMISSION*

○ Pre-application ● Application ○ Changed/Corrected Application

4.a. Federal Identifier
AG046345

b. Agency Routing Number

2. DATE SUBMITTED

Application Identifier

c. Previous Grants.gov Tracking Number

5. APPLICANT INFORMATION

Organizational DUNS*: 038633251

Legal Name*:	The Research Foundation for SUNY on behalf of U. at Buffalo
Department:	Sponsored Projects Services
Division:	
Street1*:	402 Crofts Hall
Street2:	
City*:	Buffalo
County:	Erie
State*:	NY: New York
Province:	
Country*:	USA: UNITED STATES
ZIP / Postal Code*:	14260-7016

Person to be contacted on matters involving this application

Prefix: First Name*: Bradley Middle Name: Last Name*: Bermudez Suffix:

Position/Title:	Senior Agreement Administrator
Street1*:	402 Crofts Hall
Street2:	
City*:	Buffalo
County:	Erie
State*:	NY: New York
Province:	
Country*:	USA: UNITED STATES
ZIP / Postal Code*:	14260-7016

Phone Number*: (716) 645-4383 Fax Number: (716) 645-2760 Email: bradley.bermudez@buffalo.edu

6. EMPLOYER IDENTIFICATION NUMBER *(EIN) or (TIN)** 1-146013200-F6

7. TYPE OF APPLICANT* X: Other (specify)

Other (Specify): Private Non-Profit

Small Business Organization Type ○ Women Owned ○ Socially and Economically Disadvantaged

8. TYPE OF APPLICATION*

○ New ● Resubmission

○ Renewal ○ Continuation ○ Revision

If Revision, mark appropriate box(es).

○ A. Increase Award ○ B. Decrease Award ○ C. Increase Duration

○ D. Decrease Duration ○ E. Other *(specify)* :

Is this application being submitted to other agencies?* ○Yes ●No What other Agencies?

9. NAME OF FEDERAL AGENCY*
National Institutes of Health

10. CATALOG OF FEDERAL DOMESTIC ASSISTANCE NUMBER
TITLE:

11. DESCRIPTIVE TITLE OF APPLICANT'S PROJECT*
Narrowing the gap in the genetic architecture of Alzheimer's disease

12. PROPOSED PROJECT

Start Date*	Ending Date*
07/01/2014	06/30/2019

13. CONGRESSIONAL DISTRICTS OF APPLICANT
NY-026

SF 424 (R&R) APPLICATION FOR FEDERAL ASSISTANCE Page 2

14. PROJECT DIRECTOR/PRINCIPAL INVESTIGATOR CONTACT INFORMATION

Prefix: First Name*: Kinga Middle Name: Last Name*: Szigeti Suffix:

Position/Title: Assistant Professor

Organization Name*: State University of New York at Buffalo

Department: Neurology

Division:

Street1*: 100 High Street

Street2:

City*: Buffalo

County: Erie

State*: NY: New York

Province:

Country*: USA: UNITED STATES

ZIP / Postal Code*: 14203-1126

Phone Number*: 716-859-3484 Fax Number: Email*: szigeti@buffalo.edu

15. ESTIMATED PROJECT FUNDING

a. Total Federal Funds Requested*	$1,959,831.00
b. Total Non-Federal Funds*	$0.00
c. Total Federal & Non-Federal Funds*	$1,959,831.00
d. Estimated Program Income*	$0.00

16.IS APPLICATION SUBJECT TO REVIEW BY STATE EXECUTIVE ORDER 12372 PROCESS?*

a. YES ○ THIS PREAPPLICATION/APPLICATION WAS MADE AVAILABLE TO THE STATE EXECUTIVE ORDER 12372 PROCESS FOR REVIEW ON:

DATE:

b. NO ● PROGRAM IS NOT COVERED BY E.O. 12372; OR

○ PROGRAM HAS NOT BEEN SELECTED BY STATE FOR REVIEW

17. By signing this application, I certify (1) to the statements contained in the list of certifications* and (2) that the statements herein are true, complete and accurate to the best of my knowledge. I also provide the required assurances * and agree to comply with any resulting terms if I accept an award. I am aware that any false, fictitious, or fraudulent statements or claims may subject me to criminal, civil, or administrative penalties. (U.S. Code, Title 18, Section 1001)

● I agree*

** The list of certifications and assurances, or an Internet site where you may obtain this list, is contained in the announcement or agency specific instructions.*

18. SFLLL or OTHER EXPLANATORY DOCUMENTATION File Name:

19. AUTHORIZED REPRESENTATIVE

Prefix: First Name*: Amy Middle Name: Last Name*: Lagowski Suffix:

Position/Title*: Proposal Award Coordinator

Organization Name*: The Research Foundation for SUNY on behalf of U. at Buffalo

Department: Sponsored Projects Services

Division:

Street1*: 402 Crofts Hall

Street2:

City*: Buffalo

County: Erie

State*: NY: New York

Province:

Country*: USA: UNITED STATES

ZIP / Postal Code*: 14260-7016

Phone Number*: (716) 645-4419 Fax Number: (716) 645-2760 Email*: amy.lagowski@buffalo.edu

Signature of Authorized Representative*

Amy Lagowski

Date Signed*

11/05/2013

20. PRE-APPLICATION File Name:

21. COVER LETTER ATTACHMENT File Name:1234-Cover_letter.pdf

424 R&R and PHS-398 Specific
Table Of Contents

Page Numbers

OMB Number: 4040-0010
Expiration Date: 06/30/2016

Project/Performance Site Location(s)

Project/Performance Site Primary Location

❍ I am submitting an application as an individual, and not on behalf of a company, state, local or tribal government, academia, or other type of organization.

Organization Name: UB, Clinical Translational Research Center
Duns Number: 0386332510000
Street1*: 875 Ellicott Street
Street2:
City*: Buffalo
County: Erie
State*: NY: New York
Province:
Country*: USA: UNITED STATES
Zip / Postal Code*: 14203-1126
Project/Performance Site Congressional District*: NY-026

Project/Performance Site Location 1

❍ I am submitting an application as an individual, and not on behalf of a company, state, local or tribal government, academia, or other type of organization.

Organization Name: Roswell Park Cancer Institute
DUNS Number: 8247710340000
Street1*: Elm & Carlton Streets
Street2:
City*: Buffalo
County:
State*: NY: New York
Province:
Country*: USA: UNITED STATES
Zip / Postal Code*: 14263-0001
Project/Performance Site Congressional District*: NY-026

File Name

Additional Location(s)

OMB Number: 4040-0001
Expiration Date: 06/30/2016

RESEARCH & RELATED Other Project Information

1. Are Human Subjects Involved?* ● **Yes** ○ **No**

1.a. If YES to Human Subjects

Is the Project Exempt from Federal regulations? ● Yes ○ No

If YES, check appropriate exemption number: _ 1 _ 2 _ 3 ✓ 4 _ 5 _ 6

If NO, is the IRB review Pending? ○ **Yes** ○ **No**

IRB Approval Date:

Human Subject Assurance Number 00008824

2. Are Vertebrate Animals Used?* ○ **Yes** ● **No**

2.a. If YES to Vertebrate Animals

Is the IACUC review Pending? ○ Yes ○ No

IACUC Approval Date:

Animal Welfare Assurance Number

3. Is proprietary/privileged information included in the application?* ○ **Yes** ● **No**

4.a. Does this project have an actual or potential impact - positive or negative - on the environment?* ○ **Yes** ● **No**

4.b. If yes, please explain:

4.c. If this project has an actual or potential impact on the environment, has an exemption been authorized or an environmental assessment (EA) or environmental impact statement (EIS) been performed? ○ Yes ○ No

4.d. If yes, please explain:

5. Is the research performance site designated, or eligible to be designated, as a historic place?* ○ **Yes** ● **No**

5.a. If yes, please explain:

6. Does this project involve activities outside the United States or partnership with international collaborators?* ○ **Yes** ● **No**

6.a. If yes, identify countries:

6.b. Optional Explanation:

	Filename
7. Project Summary/Abstract*	1235-PROJECT SUMMARY.pdf
8. Project Narrative*	1236-PROJECT_NARRATIVE.pdf
9. Bibliography & References Cited	1237-Bibliography and references cited.pdf
10.Facilities & Other Resources	1238-FACILITIES.pdf
11.Equipment	1239-MAJOR_EQUIPMENT.pdf

Here we seek to understand the contribution of copy number variation to the genetic architecture of Alzheimer’s disease. Understanding the genetics of AD can contribute in various ways to the development of the field. (i) As genetic biomarkers are mostly stable, they can serve as biomarkers for risk stratification and early diagnosis in order to facilitate early intervention and success of disease modifying therapy, and (ii) genetics can delineate subsets within AD, and suggest subset specific treatment options. The failure of the numerous clinical trials over the last two decade prompts us to rethink and to try alternative approaches. GWAS studies suggest that AD is genetically heterogeneous, which likely implicates heterogeneity in disease pathogenesis as well. Genetic risk/pathomechanism specific treatment could result in success.

Between the powerful SNP-GWAS studies and the planned Whole Genome Sequencing projects we have not adequately assessed the contribution of copy number variation (CNV) to the genetic architecture of AD. CNV studies leveraging the SNP arrays used in traditional GWAS face multiple challenges and to overcome these difficulties, CNV analyses of SNP arrays in AD focused on high stringency calls. There is a need for optimally powered studies using existing GWAS data, for studies that fully explore the existing GWAS data, and studies applying high density aCGH to evaluate small CNVs.

We propose to **1) refine** the contribution of CNVs to the genetic architecture of AD by performing a CNV GWAS using the Alzheimer’s Disease Genetic Consortium dataset. By optimizing normalization methods and logR ratio calculations, performing segmentation only to reduce the dataset where events may occur, performing the test of association on the numeric segmented data, the resolution and power is markedly improved. The segmentation without calls will allow the detection of smaller events, limited by the genomic coverage of the SNP array used. **2) elucidate** the contribution of small CNVs to the genetic architecture of AD by performing a CNV-eQTL association study on 200 AD and 200 normal control temporal lobes using custom Agilent array and Illumina RNA-seq. Existing datasets do not have the resolution to detect CNVs in the 150 bp-50 kb range and WGS will have limitations to assemble events in this size range due to the short reads, creating a gap in the size of CNVs assessed. In a genetically heterogeneous disorder, utilizing eQTL and CNVs as a genetic marker map within the same individuals in the context of case control status is a robust method to elucidate meaningful associations.

Alzheimer’s disease is the most common form of dementia affecting 5.4 million people in the US. The disease has a large genetic component, thus understanding the genetics of Alzheimer’s disease can help with diagnosing the disease early and may suggest what type of medications would work in some or all patients.

FACILITIES & OTHER RESOURCES – University at Buffalo, SUNY

Environment – Contribution to Success

The facilities and other resources available to the PI at the Clinical and Translational Research Center (CTRC) include everything needed to undertake and complete the proposed research project successfully. When she set up her office and adjacent laboratory they were equipped with this project specifically in mind. The intellectual environment is rich with other extramurally funded investigators who are doing work that is complementary to what is proposed here. These facilities, together with those described for the other project/performance sites, collectively provide a scientific environment that is strongly supportive of the proposed research and, therefore, success of the project.

Institutional Commitment to ESI

The PI qualifies as an Early Stage Investigator. There is extensive evidence of institutional commitment to her development as an academic researcher. Her tenure-track academic appointment includes 81% effort dedicated to research. The start-up package provided to her included laboratory space and office in the Clinical and Translational Research Center, equipment and research funds needed to launch her program. The package was sufficient to yield the preliminary data needed for this first R01 application. Support for the technician and laboratory manager, Deepika Lal, that is requested in this application was also provided by the PI's department so that a highly qualified individual could be recruited in advance of making this proposal. Administrative support is provided to the PI by a departmental administrative core. Important career-development programs are also available to her, including a formal course on the ethical conduct of research; a workshop designed to develop proposal-writing skills; and a faculty mentoring committee for early stage investigators consisting of four members: Timothy Murphy, MD, distinguished professor, the director of the CTRC, Suzanne Laychock, PhD, Professor, Senior Associate Dean and Susan Udin, PhD and Susan Baker, MD, PhD. The PI's research area is aligned to UB's strategic strength of "Health and Wellness through the Lifespan", which will help to ensure continued institutional commitment to his program.
Facilities:

Laboratory: The PI is assigned a 1200 sq ft laboratory that is located in departmental space, adjacent to her office. It is subdivided into a general purpose area (800 sq ft), a tissue culture room (200 sq ft) and a walk-in cold room. The Lab is fully equipped for all the proposed experiments and includes an Agilent scanner. The PI has access to the UB genomics core facility that provides RNA sequencing and the iPSC core.

Clinical:

The PI directs the Alzheimer's Disease and Memory Disorders Center with clinical service one day a week.

Computer:

The PI has access the computing facilities at the Center of Excellence for Bioinformatics at UB and her own lab is equipped with multiple computers for aCGH and sequencing analyses which are ongoing. The Center of Excellence for Bioinformatics at UB is a leading academic supercomputing facility. The Center for Computational Research (CCR) at COE has more than 100 Tflops of peak performance compute capacity and 600 TB of high-performance storage. CCR's computing facilities, which are housed in a state-of-the-art 4000 sq ft machine room, include a Linux cluster with more than 8000 processor cores and QDR Infiniband interconnect. A subset (32) of the cluster nodes contain (64) NVidia Tesla M2050 "Fermi" graphics processing units (GPUs). The Center also maintains several high-performance storage systems including Isilon-based storage (320TB) as well as a parallel storage system from Panasas (170TB). The computer visualization laboratory features a tiled display wall, and a VisDuo passive stereo system.

Consortium with Roswell Park Cancer Institute

The Bioinformatics Shared Resource supports the research needs of Roswell Park Cancer Institute (RPCI) investigators with respect to experimental design, data integration, as well as data analysis of clinical, laboratory, and population-based studies utilizing high-throughput bioinformatics technologies. Our expertise lies in developing and utilizing state-of-the-art bioinformatics algorithms to design, analyze and integrate various voluminous "-omics" data for a better understanding of the molecular mechanisms underlying human cancer initiation, progression and prognosis. Projects benefit from the interactive, collaborative and supportive environment.

Biostatistics Resource: The Biostatistics Resource ensures that biostatistical, bioinformatics and biomathematical support is readily available to basic, clinical and population-oriented RPCI collaborators. Services include exploratory data analysis; grant writing (statistical methods sections); hypothesis formulation; experimental design; fitting models to data; simulating data from models; developing customized data mining software and developing novel methods based on emerging technologies (e.g. cutting-edge microarray platforms, computer-intensive statistical methods).   The Bioinformatics Resource is a sub-component of the Biostatistics Resource. Bioinformatics staff has experience and expertise in bioinformatics software/database development, customized data analysis, and providing bioinformatics infrastructure and training. Services include, but are not limited to: development of bioinformatics software and databases, integrative analysis of multi-dimensional high-throughput data (e.g. gene expression, array CGH, CHIP-chip, SNPs, DNA sequencing data, etc.), identification of transcriptional regulatory elements, mapping of pathways and gene ontology, gene annotation, identification of alternative splicing forms, etc.

MAJOR EQUIPMENT

The PI has an Agilent hybridization oven, scanner, centrifuges and microcentrifuges, nanodrop and electrophoresis systems, a tissue culture room with hoods and incubators within the Clinical Translational Research Center at the University at Buffalo. Through the CTRC core equipment the PI has access to multiple slow speed and ultracentrifuges, cold rooms, a warm room, a darkroom, additional thermal cyclers and a spectrophotometer.

RESEARCH & RELATED Senior/Key Person Profile (Expanded)

PROFILE - Project Director/Principal Investigator

Prefix: First Name*: Kinga Middle Name Last Name*: Szigeti Suffix:

Position/Title*: Assistant Professor
Organization Name*: State University of New York at Buffalo
Department: Neurology
Division:
Street1*: 100 High Street
Street2:
City*: Buffalo
County: Erie
State*: NY: New York
Province:
Country*: USA: UNITED STATES
Zip / Postal Code*: 14203-1126

Phone Number*: 716-859-3484 Fax Number: E-Mail*: szigeti@buffalo.edu

Credential, e.g., agency login: kingas

Project Role*: PD/PI Other Project Role Category:

Degree Type: PhD Degree Year: 2006

File Name

Attach Biographical Sketch*: 1243-BIOGRAPHICAL SKETCH.pdf

Attach Current & Pending Support:

PROFILE - Senior/Key Person

Prefix: First Name*: Li Middle Name Last Name*: Yan Suffix: Ph.D

Position/Title*: Senior Biostatistician
Organization Name*: Roswell Park Cancer Institute
Department: Bioinformatics
Division:
Street1*: Elm & Carlton Streets
Street2:
City*: Buffalo
County: Erie
State*: NY: New York
Province:
Country*: USA: UNITED STATES
Zip / Postal Code*: 14263-0001

Phone Number*: 716-845-7757 Fax Number: E-Mail*: liyan@buffalo.edu

Credential, e.g., agency login:

Project Role*: Other (Specify) Other Project Role Category: Senior Biostatistician

Degree Type: PhD Degree Year:

File Name

Attach Biographical Sketch*: 1244-YanBiosketch.pdf

Attach Current & Pending Support:

PROFILE - Senior/Key Person				
Prefix:	First Name*: Jianmin	Middle Name	Last Name*: Wang	Suffix: Ph.D

Position/Title*:	Assistant Member, Clinical Research
Organization Name*:	Roswell Park Cancer Institute
Department:	Bioinformatics
Division:	
Street1*:	Elm & Carlton Streets
Street2:	
City*:	Buffalo
County:	Erie
State*:	NY: New York
Province:	
Country*:	USA: UNITED STATES
Zip / Postal Code*:	14263-0001

Phone Number*: 716-845-1499 Fax Number: E-Mail*: jianmin.wang@roswellpark.org

Credential, e.g., agency login:

Project Role*: Co-Investigator	Other Project Role Category:
Degree Type: PhD	Degree Year:

	File Name
Attach Biographical Sketch*:	1245-WangBiosketch.pdf
Attach Current & Pending Support:	

BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FOUR PAGES.**

NAME Szigeti, Kinga	POSITION TITLE Assistant Professor of Neurology
eRA COMMONS USER NAME kingas	

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable.)*

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	MM/YY	FIELD OF STUDY
University of Pecs Medical School, Hungary	M.D.	09/94	Medicine
Harvard University	Postdoctoral fellow	09/97-7/98	Molecular Biology
SUNY University of Buffalo	Residency	7/98-6/02	Neurology
Baylor College of Medicine	Fellowship	7/02-6/04	Molecular and Human Genetics
University of Szeged Medical School, Hungary	Ph.D.	05/06	Clinical Neuroscience

A. Personal Statement

The proposed research will take place under the auspices of the Clinical and Translational Research Center at the University at Buffalo, The Center of Excellence for Bioinformatics at UB and the Alzheimer's Disease and Memory Disorders Center. I am a board certified neurologist with subspecialty training in genetics (Baylor College of Medicine) and the founding director of the ADMDC at Buffalo. I have expertise in cognitive disorders clinical care and genetic research relevant to common diseases, including genome-wide association studies. Our research focus is to understand the contribution of copy number variation to the genetic architecture of Alzheimer's disease. Understanding the genetics of AD can contribute in various ways to the development of the field. (i) As genetic biomarkers are mostly stable, they can serve as biomarkers for risk stratification and early diagnosis in order to facilitate early intervention and success of disease modifying therapy, and (ii) genetics can delineate subsets within AD, and suggest subset specific treatment options. The failure of the numerous clinical trials over the last two decade prompts us to rethink and to try alternative approaches. GWAS studies suggest that AD is genetically heterogeneous, which likely implicates heterogeneity in disease pathogenesis as well. Genetic risk/pathomechanism specific treatment could result in success. Through an unbiased approach (CNV-GWAS) we seek to identify pathomechanisms that may pertain to a subset of patients while also developing high-throughput cell based assays (iPSC) for drug screens with the ultimate goal to develop novel treatment strategies.

B. Positions and Honors

<u>Positions and Employment</u>

1994-1997 Postdoctoral Fellow, Molecular Biology, University of Pecs, Hungary
1997-1998 Postdoctoral Fellow, Molecular Pathology, Harvard University
1998-2002 Medical Residency, Neurology, State University of New York at Buffalo
2002-2004 Fellowship, Molecular and Human Genetics, Baylor College of Medicine
2005-2006 Staff Neurologist, United Health System, Hungary
2006-2010 Assistant Professor of Neurology and Molecular and Human Genetics, Alzheimer Disease and Memory Disorders Center, Baylor College of Medicine

2010-present Assistant Professor of Neurology, SUNY Buffalo

Other Experience and Professional Memberships

2001-2002 Chief Resident
2002-present Member, American Society of Human Genetics
2003 Diplomate, American Board of Neurology and Psychiatry
2004 Contributor, Polyneuropathy Task Force of the American Academy of Neurology
2005-2006 Clinical Research Scientist, Richter Gedeon, Ltd, Budapest, Hungary
2012 Associate Editor, Journal of Alzheimer's Disease
2012 Associate Editor, American Journal of Neurodegeneration

Honors

2002 Recipient of the Michael E. Cohen Residency Research Award

C. Selected Peer-reviewed Publications

Most relevant to the current application

1. Shaw, C.A., Y. Li, J. Wiszniewska, S. Chasse, S.N. Zaidi, W. Jin, B. Dawson, K. Wilhelmsen, J.R. Lupski, J.W. Belmont, R.S. Doody, and **K. Szigeti**, *Olfactory copy number association with age at onset of Alzheimer disease.* Neurology, 2011. **76**(15): p. 1302-9. PMCID: PMC3090061
2. Li, Y., C.A. Shaw, I. Sheffer, N. Sule, S.Z. Powell, B. Dawson, S.N. Zaidi, K.L. Bucasas, J.R. Lupski, K.C. Wilhelmsen, R. Doody, and **K. Szigeti**, *Integrated copy number and gene expression analysis detects a CREB1 association with Alzheimer's disease.* Transl Psychiatry, 2012. **2**: p. e192. PMID: 23168992
3. **Szigeti, K**., D. Lal, Y. Li, R.S. Doody, K. Wilhelmsen, L. Yan, S. Liu, and C. Ma, *Genome-wide scan for copy number variation association with age at onset of Alzheimer's disease.* J Alzheimers Dis, 2013. **33**(2): p. 517-23. PMID: 23202439

Additional recent publications of importance to the field (in chronological order)

4. Saifi, G.M., **K. Szigeti**, G.J. Snipes, C.A. Garcia, and J.R. Lupski, *Molecular mechanisms, diagnosis, and rational approaches to management of and therapy for Charcot-Marie-Tooth disease and related peripheral neuropathies.* J Investig Med, 2003. **51**(5): p. 261-83. PMID:14577517
5. **Szigeti, K.**, G.M. Saifi, D. Armstrong, J.W. Belmont, G. Miller, and J.R. Lupski, *Disturbance of muscle fiber differentiation in congenital hypomyelinating neuropathy caused by a novel myelin protein zero mutation.* Ann Neurol, 2003. **54**(3): p. 398-402. PMID: 12953275
6. **Szigeti, K**., N. Sule, A.M. Adesina, D.L. Armstrong, G.M. Saifi, E. Bonilla, M. Hirano, and J.R. Lupski, *Increased blood-brain barrier permeability with thymidine phosphorylase deficiency.* Ann Neurol, 2004. **56**(6): p. 881-6. PMID: 15562405
7. **Szigeti, K**., L.J. Wong, C.L. Perng, G.M. Saifi, K. Eldin, A.M. Adesina, D.L. Cass, M. Hirano, J.R. Lupski, and F. Scaglia, *MNGIE with lack of skeletal muscle involvement and a novel TP splice site mutation.* J Med Genet, 2004. **41**(2): p. 125-9. PMCID: PMC1735672
8. Saifi, G.M., **K. Szigeti**, W. Wiszniewski, M.E. Shy, K. Krajewski, I. Hausmanowa-Petrusewicz, A. Kochanski, S. Reeser, P. Mancias, I. Butler, and J.R. Lupski, *SIMPLE mutations in Charcot-Marie-Tooth disease and the potential role of its protein product in protein degradation.* Hum Mutat, 2005. **25**(4): p. 372-83. PMID: 15776429
9. **Szigeti, K**., C.A. Garcia, and J.R. Lupski, *Charcot-Marie-Tooth disease and related hereditary polyneuropathies: molecular diagnostics determine aspects of medical management.* Genet Med, 2006. **8**(2): p. 86-92. PMID:16481890
10. Verhoeven, K., K.G. Claeys, S. Zuchner, J.M. Schroder, J. Weis, C. Ceuterick, A. Jordanova, E. Nelis, E. De Vriendt, M. Van Hul, P. Seeman, R. Mazanec, G.M. Saifi, **K. Szigeti**, P. Mancias, I.J. Butler, A. Kochanski, B. Ryniewicz, J. De Bleecker, P. Van den Bergh, C. Verellen, R. Van Coster, N. Goemans,

M. Auer-Grumbach, W. Robberecht, V. Milic Rasic, Y. Nevo, I. Tournev, V. Guergueltcheva, F. Roelens, P. Vieregge, P. Vinci, M.T. Moreno, H.J. Christen, M.E. Shy, J.R. Lupski, J.M. Vance, P. De Jonghe, and V. Timmerman, *MFN2 mutation distribution and genotype/phenotype correlation in Charcot-Marie-Tooth type 2.* Brain, 2006. **129**(Pt 8): p. 2093-102. PMID: 16714318

11. Chow, C.Y., Y. Zhang, J.J. Dowling, N. Jin, M. Adamska, K. Shiga, **K. Szigeti**, M.E. Shy, J. Li, X. Zhang, J.R. Lupski, L.S. Weisman, and M.H. Meisler, *Mutation of FIG4 causes neurodegeneration in the pale tremor mouse and patients with CMT4J.* Nature, 2007. **448**(7149): p. 68-72. PMCID: PMC2271033
12. **Szigeti, K**., W. Wiszniewski, G.M. Saifi, D.L. Sherman, N. Sule, A.M. Adesina, P. Mancias, S. Papasozomenos, G. Miller, L. Keppen, D. Daentl, P.J. Brophy, and J.R. Lupski, *Functional, histopathologic and natural history study of neuropathy associated with EGR2 mutations.* Neurogenetics, 2007. **8**(4): p. 257-62. PMID: 17717711
13. England, J.D., G.S. Gronseth, G. Franklin, G.T. Carter, L.J. Kinsella, J.A. Cohen, A.K. Asbury, **K. Szigeti**, J.R. Lupski, N. Latov, R.A. Lewis, P.A. Low, M.A. Fisher, D.N. Herrmann, J.F. Howard, Jr., G. Lauria, R.G. Miller, M. Polydefkis, and A.J. Sumner, *Practice Parameter: evaluation of distal symmetric polyneuropathy: role of laboratory and genetic testing (an evidence-based review). Report of the American Academy of Neurology, American Association of Neuromuscular and Electrodiagnostic Medicine, and American Academy of Physical Medicine and Rehabilitation.* Neurology, 2009. **72**(2): p. 185-92. PMID: 19056666
14. McLaughlin, H.M., R. Sakaguchi, C. Liu, T. Igarashi, D. Pehlivan, K. Chu, R. Iyer, P. Cruz, P.F. Cherukuri, N.F. Hansen, J.C. Mullikin, L.G. Biesecker, T.E. Wilson, V. Ionasescu, G. Nicholson, C. Searby, K. Talbot, J.M. Vance, S. Zuchner, **K. Szigeti**, J.R. Lupski, Y.M. Hou, E.D. Green, and A. Antonellis, *Compound heterozygosity for loss-of-function lysyl-tRNA synthetase mutations in a patient with peripheral neuropathy.* Am J Hum Genet, 2010. **87**(4): p. 560-6. PMCID: PMC2948804
15. **Szigeti, K**. and R.S. Doody, *Should persons with autosomal dominant AD be included in clinical trials? Authors' response.* Alzheimers Res Ther, 2011. **3**(3): p. 19. PMCID: PMC3226308

D. Research Support

Ongoing Research Support

NIA K23 AG036852 Szigeti (PI) 10/2010-09/2015

Project Title: Olfactory receptor copy number association with age at onset of Alzheimer disease
We propose to develop a custom aCGH covering all OR genes to comprehensively study the ORs. The array approach allows complete coverage of all of the OR regions in order to establish their modifier effect on AAO. This study will be an add-on to an ongoing multicenter prospective cohort study in the State of Texas (Texas Alzheimer Research Consortium, TARC). This aspect of the proposal will provide the preliminary data for future clinical studies for example addressing the predictive value of the modifiers in prospective cohorts of patients with MCI. 850 subjects (500 AD, 250 MCI and 100 normal controls) will be enrolled. Baseline and yearly longitudinal neuropsychological measures are being ascertained and entered into the TARC database. We will correlate the olfactory genotype and phenotype using the aCGH and the University of Pennsylvania Smell Identification Test (UPSIT) measure. We will compare cognitive profiles at presentation between the various OR copy number states to assess whether there is a cognitive endophenotype associated with the OR genotypes.

SUNY Buffalo Szigeti (PI) 09/2010-08/2013
Start-up funds

Project Title: Longitudinal genetic study in Alzheimer's disease and amnestic mild cognitive impairment. Clinical, neuropsychological, MRI data and DNA is collected in the longitudinal cohort at the Alzheimer's Disease and Memory Disorders Center with the goal to establish a pool of subjects with interest in genetic studies. Candidates found in the GWAS datasets are tested and deep-phenotyped in this cohort, including iPSC in some cases. Subjects were asked for permission to be recontacted for further studies.

Completed Research Support

Alzheimer Association NIRG-10-174455 Szigeti (PI) 07/2010-06/2012

Project Title: Copy number variation GWA with age at onset of Alzheimer disease
A genome-wide association study to identify loci that affect age at onset of Alzheimer disease.
As part of the Texas Alzheimer Research Consortium we are conducting a SNP genome-wide association (GWA) with age at onset of Alzheimer Disease study on the Genome-Wide Human SNP Array 6.0 (Affymetrix) (500 patients and 100 controls). This platform interrogates copy number variation (CNVs) by the addition of more than 946,000 copy number probes. We completed CNV genome wide association (GWA) with AAO using hazard regression model and ordered subset analysis.

BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors in the order listed on Form Page 2.
Follow this format for each person. **DO NOT EXCEED FOUR PAGES.**

NAME Yan, Li	POSITION TITLE Senior Statistician
eRA COMMONS USER NAME (credential, e.g., agency login) YAN_LI	

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable.)*

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	MM/YY	FIELD OF STUDY
Shandong University, Shandong, China	BS	1988	Physics
Shandong University, Shandong, China	MS	1993	Physics
University of Rochester, NY	PhD	2002	Physics
University at Buffalo, SUNY	MA	2009	Biostatistics

A. Personal Statement:

As a senior biostatistician at Roswell Park Cancer Institute, I collaborated with Dr. Szigeti (PI of this proposal) on several projects using cutting-edge statistics and genomics method to study the risk factors of Alzheimer disease. I would provide statistical support on experimental design, clinical data and genomic data integration, and analysis the integrated data for the specified aims. I have the expertise and motivation necessary to successfully carry out the proposed work. With a broad background in bioinformatics, biostatistics, and population, clinical related statistical support, I have accumulated extensive experience in statistical and bioinformatics analysis, incorporating the conventional clinical data analysis and high-throughput genomics datasets including large-scale high-density microarray and second-generation sequencing studies.

B. Positions and Honors

Positions and Employment

2002-2004	Statistical Research Coordinator, Department of Biostatistics, University of Florida, Gainesville, FL
2004-2005	Assistant Research Scientist, Department of Biostatistics, University of Florida, Gainesville, FL
2005-2006	Research Associates, Frontier Science and Technology Research Foundation, Buffalo, NY
2006-2010	Research Assistant Professor, Department of Biostatistics, University at Buffalo, Buffalo, NY
2010-present	Senior statistician, Department of Biostatistics, Roswell Park Cancer Institute, Buffalo, NY

Honors

1998-2002	Kodak Fellowship

C. Selected peer-reviewed publications or manuscripts in press

Most Relevant to the current application

1. Tian, L., Li, X., **Yan, L**. Testing equality of generalized treatment effects. J Biopharm Stat Vol.22, 582 (2012).
2. Shen J, Wang D,Gregory S, Medico L, Hu Q, **Yan L**, Odunsi K, Lele S B, Ambrosone C, Liu S, Zhao H. (2012) Evaluation of microRNA expression profiles and their associations with risk alleles in lymphoblastoid cell lines of familial ovarian cancer *Carcinogenesis* Vol. 33, 604(2012).

Additional recent publications of importance to the field (in chronological order)

3. L.Tian, A.Vexler, **L.Yan,** and E. Schisterman, (2009) Confidence Interval Estimation of the Difference Between Paired AUCs Based on Combined Biomarkers', JSPI, Vol.139, 3725(2009).

4. Hu Q, Wang D, **Yan L**, Zhao H and Liu S (2012) VPA: an R package for analyzing sequencing variants with user-specified frequency pattern. *BMC Research Notes Vol. 5, 31(2012).*
5. **Yan L**, Ma C, Wang D, Hu Q, Qin M, Conroy J , Sucheston L, et al. "OSAT: a Tool for Sample-to-batch Allocations in Genomics Experiments." BMC Genomics 13, no. 1 (2012): 689

D. Research Support

Ongoing Research Support

1R01 HL102278 Hahn/Sucheston (PIs) 07/01/10–06/30/14
NHLBI
Genetic susceptibility to unrelated donor stem cell transplant-related mortality
This project will study genetic causes for unrelated donor blood or marrow transplant (BMT)-related mortality and provide initial support for a more tailored, individualized approach to selecting which chemotherapy drugs to use and at what dose.
Role on Project: Statistician

5R01 CA139426-03 Ambrosone 03/05/10-12/31/13
NHI
Genome Wide Predictors of Treatment-Related Toxicities
We propose to conduct a genome-wide scan, comparing breast cancer patients who experience severe toxicities to those who do not, in the context of a clinical trial of differing doses of doxirubicin (A), cyclophosphamide (C) and paclitaxel (T). We expect that this study will reveal genetic variants that are associated with sensitivity of normal cells to damage from chemotherapeutic agents.
Role on Project: Statistician

Recently Completed Research Support

Contract #HHSN261200800001E Zhao (PI) 08/12/2010–7/31/2012
NIH/NCI, SAIC-Frederick's
Effects of pre-analytic variables on circulating microRNAs using a CCSG biorepository
Goal is to study how the pre-analytic variables might affect the levels of circulating microRNAs as a biomarker for cancer prediction.
Role on Project: Bio-statistician

(Carter, PI) 01/01/06 – 12/31/10 (renewable)
New York State Department of Health
Child and Family Outcomes Evaluation for State Performance Plan
This project will evaluate the NYS Performance Plan for the Federal Early Childhood Outcome Assessment Project.
Role on Project: Data manager and statistician

BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors in the order listed on Form Page 2.
Follow this format for each person. **DO NOT EXCEED FOUR PAGES.**

NAME Jianmin Wang	POSITION TITLE CCSG Bioinformatics Shared Resource Co-Director, Assistant Professor of Oncology
eRA COMMONS USER NAME (credential, e.g., agency login) JIANMINWANG	

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable.)*

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	MM/YY	FIELD OF STUDY
Peking University, Beijing, China	B.S.	1997	Molecular Biology
Peking University, Beijing, China	M.S.	2000	Bioinformatics
Iowa State University, Ames, IA	Ph.D.	2006	Bioinformatics

A. Personal Statement

My primary research interests are focusing on the analysis of large scale omics data from cutting edge biotechnologies using computational and statistical methods. My research expertises include i) Sensitive identification of structure variation from next generation sequencing data. ii) Statistical inference for isoform identification and abundance estimation from RNASeq data. iii) Integrated study of genomic, epigenetic and proteomic data. I also serve as a regular reviewer for journals such as Bioinformatics.

I have been actively involved in the development of several software packages that are widely used in the sequencing community: i) PCAP, a parallel whole gene assembly program. ii) PolyFreq, a program to identify SNPs with allele frequency. iii) ChiP-PAM, a program for ChiP-Seq data analysis using motif finding. iv) CREST, a structure variation detection program for whole genome sequencing data. v) FUSIM, a simulation tool for fusion discovery from transcriptome sequencing data.

B. Positions and Honors

Positions and Employment

2006-2006	Bioinformatics Postdoctoral Associate, Roswell Park Cancer Institute, Buffalo, NY
2008-2012	Bioinformatics Scientist, St. Jude Children's Research Hospital, Memphis, TN
2012-present	Research Assistant Professor, Department of Biostatistics, SUNY at Buffalo, Buffalo, NY
2012-present	Assistant Professor, Department of Biostatistics and Bioinformatics, Roswell Park Cancer Institute, Buffalo, NY
2012-present	Co-Director, CCSG Bioinformatics Shared Resource, Roswell Park Cancer Institute
2012-present	Bioinformatics Leader, Center for Personalized Medicine, Roswell Park Cancer Institute

C. Recent peer-reviewed publications

1. Marella NV, Malyavantham KS, **Wang J,** Matsui S, Liang P, Berezney R. Cytogenetic and cDNA microarray expression analysis of MCF10 human breast cancer pregression cell lines. Cancer Res 2009; 69(14):5946-5953. PMCID:PMC2826242.
2. Kasper LH, Lerach S, **Wang J**, Wu S, Jeevan T, Brindle PK. CBP/p300 double null cells reveal effect of coactivator level and diversity on CREB transactivation. EMBO J 2010; 29(21):3660-3672. PMCID:PMC2982758.
3. Wu S, **Wang J**, Zhao W, Pounds S, Cheng C. ChIP-PaM: an algorithm to identify protein-DNA interaction using ChiP-Seq data. Theor Biol Med Model 2010; 7:18. PMCID:PMC2893127.
4. Carol H, Boehm I, Reynolds CP, Kang MH, Maris JM, Morton CL, Gorlick R, Kolb EA, Keir ST, Wu J, Wozniak AE, Yang Y, Manfredi M, Ecsedy J, **Wang J**, Neale G, Houghton PJ, Smith MA, Lock RB. Efficacy and pharmacokinetic/pharmacodynamic evaluation of the Aurora kinase A inhibitor MLN8237 against

preclinical models of pediatric cancer. Cancer Chemother Pharmacol 2011; 68(5):1291-1304. PMCID:PMC3215888.

5. Cheung NK, Zhang J, Lu C, Parker M, Bahrami A, Tickoo SK, Heguy A, Pappo AS, Federico S, Dalton J, Cheung IY, Ding L, Fulton R, **Wang J**, Chen X, Becksfort J, Wu J, Billups CA, Ellison D, Mardis ER, Wilson RK, Downing JR, Dyer MA. Association of age at diagnosis and genetic mutations in patients with neuroblastoma. JAMA 2012; 307(10):1062-1071. PMCID:PMC3527076.
6. Zhang J, McEvory J, Flores-Otero J, Ding L, Chen X, Wilson M, Wu G, **Wang J**, et al. A novel retinoblastoma therapy from genomic and epigenetic analyses. Nature 2012; 481(7381):329-334. PMCID: PMC3289956
7. Zhang J, Ding L, Holmfeldt L, Wu G, Heatley SL, Payne-Turner D, Easton J, Chen X, **Wang J**, Rusch M, Lu C, Chen SC, Wei L, Collins-Underwood JR, Ma J, Roberts KG, Pounds SB, Ulyanov A, Becksfort J, Gupta P, Huether R, Kriwacki RW, Parker M, McGoldrick DJ, Zhao D, Alford D, Espy S, Bobba KC, Song G, Pei D, Cheng C, Roberts S, Barbato MI, Campana D, Coustan-Smith E, Shurtleff SA, Raimondi SC, Kleppe M, Cools J, Shimano KA, Hermiston ML, Doulatov S, Eppert K, Laurenti E, Notta F, Dick JE, Basso G, Hunger SP, Loh ML, Devidas M, Wood B, Winter S, Dunsmore KP, Fulton RS, Fulton LL, Hong X, Harris CC, Dooling DJ, Ochoa K, Johnson KJ, Obenauer JC, Evans WE, Pui CH, Naeve CW, Ley TJ, Mardis ER, Wilson RK, Downing JR, Mullighan CG. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. Nature 2012; 481(7380):157-63. PMCID:PMC3267575.
8. **Wang J**, Mullighan CG, Easton J, Roberts S, Heatley SL, Ma J, Rusch MC, Chen K, Harris CC, Ding L, Holmfeldt L, Payne-Turner D, Fan X, Wei L, Zhao D, Obenauer JC, Naeve C, Mardis ER, Wilson RK, Downing JR, Zhang J. CREST: an algorithm that maps structure variation with base-pair resolution. Nat Methods 2011; 8(8):652-654. PMCID:PMC3527068
9. Robinson G, Parker M, Kranenburg TA, Lu C, Chen X, Ding L, Phoenix TN, Hedlund E, Wei L, Zhu X, Chalhoub N, Baker SJ, Huether R, Kriwacki R, Curley N, Thiruvenkatam R, **Wang J**, Wu G, Rusch M, Hong X, Becksfort J, Gupta P, Ma J, Easton J, Vadodaria B, Onar-Thomas A, Lin T, Li S, Pounds S, Paugh S, Zhao D, Kawauchi D, Roussel MF, Finkelstein D, Ellison DW, Lau CC, Bouffet E, Hassall T, Gururangan S, Cohn R, Fulton RS, Fulton LL, Dooling DJ, Ochoa K, Gajjar A, Mardis ER, Wilson RK, Downing JR, Zhang J, Gilbertson RJ. Novel mutations target distinct subgroups of medulloblastoma. Nature 2012; 488(7409):43-48. PMCID:PMC3412905.
10. Gruber TA, Larson Gedman A, Zhang J, Koss CS, Marada S, Ta HQ, Chen SC, Su X, Ogden SK, Dang J, Wu G, Gupta V, Andersson AK, Pounds S, Shi L, Easton J, Barbato MI, Mulder HL, Manne J**, Wang J**, Rusch M, Ranade S, Ganti R, Parker M, Ma J, Radtke I, Ding L, Cazzaniga G, Biondi A, Kornblau SM, Ravandi F, Kantarjian H, Nimer SD, Döhner K, Döhner H, Ley TJ, Ballerini P, Shurtleff S, Tomizawa D, Adachi S, Hayashi Y, Tawa A, Shih LY, Liang DC, Rubnitz JE, Pui CH, Mardis ER, Wilson RK, Downing JR. An Inv(16)(p13.3q24.3)-encoded CBFA2T3-GLIS2 fusion protein defines an aggressive subtype of pediatric acute megakaryoblastic leukemia. Cancer Cell 2012(22):683-697.
11. Yan L, Ma C, Wang D, Hu Q, Qin M, Conroy JM, Sucheston LE, Ambrosone CB, Johnson CS, **Wang J**, Liu S. OSAT: a tool for sample-to-batch allocations in genomics experiments. BMC Genomics 2012; 13:689. PMCID:PMC3548766
12. Bruno A, Miecznikowski J, Qian M, **Wang J**, Liu S. FUSIM: a software tool for simulating fusion transcripts. BMC Bioinformatics 2013; 14(1):13. PMID: 23323884.
13. The St. Jude Children's Research Hospital–Washington University Pediatric Cancer Genome Project, Zhang J, Wu G, Miller CP, Tatevossian RG, Dalton JD, Tang B, Orisme W, Punchihewa C, Parker M, Qaddoumi I, Boop FA, Lu C, Kandoth C, Ding L, Lee R, Huether R, Chen X, Hedlund E, Nagahawatte P, Rusch M, Boggs K, Cheng J, Becksfort J, Ma J, Song G, Li Y, Wei L, **Wang J**, Shurtleff S, Easton J, Zhao D, Fulton RS, Fulton LL, Dooling DJ, Vadodaria B, Mulder HL, Tang C, Ochoa K, Mullighan CG, Gajjar A, Kriwacki R, Sheer D, Gilbertson RJ, Mardis ER, Wilson RK, Downing JR, Baker SJ, Ellison DW. Whole-genome sequencing identifies genetic alterations in pediatric low-grade gliomas. Nature Genetics. 2013 Apr 14. doi: 10.1038/ng.2611. [Epub ahead of print] PMID: 23583981
14. Khoury T, Hu Q, Liu S, **Wang J.** Intracystic papillary carcinoma of breast: interrelationship with in situ and invasive carcinoma and a proposal of pathogenesis: array comparative genomic hybridization study of 14 cases. Mod Pathol 2013; Aug 2., PMID: 23907150

15. Baysal BE, De Jong K, Liu B, **Wang J,** Patnaik SK, Wallace PK, Taggart RT. Hypoxia-induced C-to-U coding RNA editing downregulates SDHB in monocytes. PeerJ 2013; Sep 10. PMID: 24058882.

D. Research Support

Ongoing

5P30CA016056-36 (Trump) 5/1/12-4/30/14
NIH / NCI (Developmental Funds)
Targeting the Tumor Mutatome for Personalized Vaccination Therapy in Ovarian Cancer
By sequencing exome and transcriptome to identify expressed nonsynonymous, primarily missense, mutations resulting in unique antigenic epitopes, such that we can test the ability of the predicted HLA class I epitopes to stimulate CD8+ cells derived from peripheral blood and tumors of the patients.
Role: Principal Investigator of developmental subproject

Completed

None

PHS 398 Cover Page Supplement

OMB Number: 0925-0001

1. Project Director / Principal Investigator (PD/PI)

Prefix:
First Name*: Kinga
Middle Name:
Last Name*: Szigeti
Suffix:

2. Human Subjects

Clinical Trial? ● No ○ Yes
Agency-Defined Phase III Clinical Trial?* ○ No ○ Yes

3. Permission Statement*

If this application does not result in an award, is the Government permitted to disclose the title of your proposed project, and the name, address, telephone number and e-mail address of the official signing for the applicant organization, to organizations that may be interested in contacting you for further information (e.g., possible collaborations, investment)?

● Yes ○ No

4. Program Income*

Is program income anticipated during the periods for which the grant support is requested? ○ Yes ● No

If you checked "yes" above (indicating that program income is anticipated), then use the format below to reflect the amount and source(s). Otherwise, leave this section blank.

Budget Period*	Anticipated Amount ($)*	Source(s)*

PHS 398 Cover Page Supplement

5. Human Embryonic Stem Cells

Does the proposed project involve human embryonic stem cells?* ● No ○ Yes

If the proposed project involves human embryonic stem cells, list below the registration number of the specific cell line(s) from the following list: http://grants.nih.gov/stem_cells/registry/current.htm. Or, if a specific stem cell line cannot be referenced at this time, please check the box indicating that one from the registry will be used:

Cell Line(s): Specific stem cell line cannot be referenced at this time. One from the registry will be used.

6. Inventions and Patents (For renewal applications only)

Inventions and Patents*: ○ Yes ○ No

If the answer is "Yes" then please answer the following:

Previously Reported*: ○ Yes ○ No

7. Change of Investigator / Change of Institution Questions

□ Change of principal investigator / program director

Name of former principal investigator / program director:

Prefix:

First Name*:

Middle Name:

Last Name*:

Suffix:

□ Change of Grantee Institution

Name of former institution*:

PHS 398 Modular Budget

OMB Number: 0925-0001

Budget Period: 1

Start Date: 07/01/2014 **End Date:** 06/30/2015

A. Direct Costs

	Funds Requested ($)
Direct Cost less Consortium F&A*	250,000.00
Consortium F&A	9,516.00
Total Direct Costs*	**259,516.00**

B. Indirect Costs

	Indirect Cost Type	Indirect Cost Rate (%)	Indirect Cost Base ($)	**Funds Requested ($)**
1.	MTDC	59.00	244,361.00	144,173.00
2.				
3.				
4.				

Cognizant Agency DHHS Louis Martillotti 212-264-2069
(Agency Name, POC Name and Phone Number)

Indirect Cost Rate Agreement Date 01/31/2013

Total Indirect Costs **144,173.00**

C. Total Direct and Indirect Costs (A + B)

Funds Requested ($) **403,689.00**

PHS 398 Modular Budget

Budget Period: 2

Start Date: 07/01/2015 **End Date:** 06/30/2016

A. Direct Costs

	Funds Requested ($)
Direct Cost less Consortium F&A*	250,000.00
Consortium F&A	9,801.00
Total Direct Costs*	**259,801.00**

B. Indirect Costs

	Indirect Cost Type	Indirect Cost Rate (%)	Indirect Cost Base ($)	**Funds Requested ($)**
1.	MTDC	59.50	221,141.00	131,579.00
2.				
3.				
4.				

Cognizant Agency (Agency Name, POC Name and Phone Number)	DHHS Louis Martillotti 212-264-2069		
Indirect Cost Rate Agreement Date	01/31/2013	**Total Indirect Costs**	**131,579.00**

C. Total Direct and Indirect Costs (A + B)

Funds Requested ($)	**391,380.00**

PHS 398 Modular Budget

Budget Period: 3

Start Date: 07/01/2016 **End Date:** 06/30/2017

A. Direct Costs

	Funds Requested ($)
Direct Cost less Consortium F&A*	250,000.00
Consortium F&A	10,096.00
Total Direct Costs*	**260,096.00**

B. Indirect Costs

	Indirect Cost Type	Indirect Cost Rate (%)	Indirect Cost Base ($)	Funds Requested ($)
1.	MTDC	59.50	217,213.00	129,242.00
2.				
3.				
4.				

Cognizant Agency
(Agency Name, POC Name and Phone Number) DHHS Louis Martillotti 212-264-2069

Indirect Cost Rate Agreement Date 01/31/2013

Total Indirect Costs **129,242.00**

C. Total Direct and Indirect Costs (A + B)

Funds Requested ($) **389,338.00**

PHS 398 Modular Budget

Budget Period: 4

Start Date: 07/01/2017 **End Date:** 06/30/2018

A. Direct Costs

	Funds Requested ($)
Direct Cost less Consortium F&A*	250,000.00
Consortium F&A	10,396.00
Total Direct Costs*	**260,396.00**

B. Indirect Costs

	Indirect Cost Type	Indirect Cost Rate (%)	Indirect Cost Base ($)	Funds Requested ($)
1.	MTDC	59.50	214,943.00	127,891.00
2.				
3.				
4.				

Cognizant Agency
(Agency Name, POC Name and Phone Number) DHHS Louis Martillotti 212-264-2069

Indirect Cost Rate Agreement Date 01/31/2013

Total Indirect Costs **127,891.00**

C. Total Direct and Indirect Costs (A + B)

Funds Requested ($) **388,287.00**

PHS 398 Modular Budget

Budget Period: 5

Start Date: 07/01/2018 **End Date:** 06/30/2019

A. Direct Costs

	Funds Requested ($)
Direct Cost less Consortium F&A*	250,000.00
Consortium F&A	10,711.00
Total Direct Costs*	**260,711.00**

B. Indirect Costs

	Indirect Cost Type	Indirect Cost Rate (%)	Indirect Cost Base ($)	Funds Requested ($)
1.	MTDC	59.50	212,480.00	126,426.00
2.				
3.				
4.				

Cognizant Agency DHHS Louis Martillotti 212-264-2069
(Agency Name, POC Name and Phone Number)

Indirect Cost Rate Agreement Date 01/31/2013

Total Indirect Costs **126,426.00**

C. Total Direct and Indirect Costs (A + B)

Funds Requested ($) **387,137.00**

PHS 398 Modular Budget

Cumulative Budget Information

1. Total Costs, Entire Project Period

Section A, Total Direct Cost less Consortium F&A for Entire Project Period ($)	1,250,000.00
Section A, Total Consortium F&A for Entire Project Period ($)	50,520.00
Section A, Total Direct Costs for Entire Project Period ($)	1,300,520.00
Section B, Total Indirect Costs for Entire Project Period ($)	659,311.00
Section C, Total Direct and Indirect Costs (A+B) for Entire Project Period ($)	1,959,831.00

2. Budget Justifications

Personnel Justification	1246-BUDGET_JUSTIFICATION.pdf
Consortium Justification	1247-Budget Justification ll.pdf
Additional Narrative Justification	

BUDGET JUSTIFICATION
PERSONNEL:

Kinga Szigeti, M.D., Ph.D., Principal Investigator***:** (3 CM) The Principal Investigator is a board certified neurologist with clinical fellowship training in molecular and human genetics. She has 5 years of postdoctoral research training. She has a PhD in clinical neuroscience. She holds a license to practice medicine in the State of New York. She is the founding director of the Alzheimer's Disease and Memory Disorders Center at the University at Buffalo. She will be coordinating every aspect of the project and will be responsible for primary data management. She will perform the statistical analyses in conjunction with Chad Shaw. She will develop and set up the assays (aCGH, RT-qPCR) and provide training for the technician and graduate student. She will oversee all the laboratory bench work of the research technician. This includes daily supervision of experimental work and data interpretation, as well as weekly formal laboratory group meetings.

Deepika Lal, Technician: (8 CM in year 1, 7 CM in year 2, 6 CM in years 3 and 4, and 5 CM in yr 5). She will be responsible for database management, sample preparation and the aCGH experiments.

TBA/Blanka Kellermayer, PhD student: (12 CM) The student will be trained and will participate in all the wet lab experiments with the technician, will receive training in bioinformatics and will be responsible on designing the orthogonal methods for validation with supervision.

***** Institutional appointment 0.82 FTE

Consortium Justification

Dr. Jianmin Wang, Assistant Member at the Health Research Inc. Roswell Park Cancer Institute (RPCI), will serve as co-Investigator to lead Bioinformatics analysis for five years (0.60 calendar months, 5% effort) and Dr. Li Yan, will serve as Senior Biostatistician for five years (0.60 calendar months, 5% effort) bioinformatics and statistical analyses. Drs Wang and Yan and the RPCI staff will participate in e-mail/ phone communications to discuss grant/project status during performance period.

PHS 398 Research Plan

Please attach applicable sections of the research plan, below.

OMB Number: 0925-0001

1. Introduction to Application (for RESUBMISSION or REVISION only)	1240-Introduction.pdf
2. Specific Aims	1241-SPECIFIC AIMS.pdf
3. Research Strategy*	1242-Research strategy.pdf
4. Progress Report Publication List	
Human Subjects Sections	
5. Protection of Human Subjects	1248-PROTECTION_OF_HUMAN_SUBJECTS.pdf
6. Inclusion of Women and Minorities	
7. Inclusion of Children	
Other Research Plan Sections	
8. Vertebrate Animals	
9. Select Agent Research	
10. Multiple PD/PI Leadership Plan	
11. Consortium/Contractual Arrangements	1249-CONSORTIUM.pdf
12. Letters of Support	1250-Letters_of_support.pdf
13. Resource Sharing Plan(s)	1251-RESOURCE_SHARING.pdf
Appendix (if applicable)	
14. Appendix	

We would like to thank the study section for the constructive criticisms and hope that all concerns have been addressed to satisfaction. Changes are depicted in ***bold, italicized*** typography. The concerns are grouped in five major categories:

1. Data and sample access

Intensity data access has been obtained from the ADNI, TGEN, NIA-LOAD, ADC 1-3, TARC and MIRAGE studies (last column Table 4). These provide a sample size of 7482 AD and 5333 normal controls. Additional collaborations are pending. We propose to use the Affymetrix 6.0 for Discovery I due to its highest density and better dynamic range for CNV analysis and the ADC 1-3 dataset for Discovery II as it has acceptable coverage and larger samples size. SNP arrays are not optimal for CNV analysis and often harbor high genotyping error rates. Due to these limitations and the variable coverage, we added an independent replication set (Baylor and UB DNA collection) for locus/CNV specific validation, genotyping error rate estimation and for potential replication with orthogonal methods. Letters for intensity file access and tissue availability are attached.

2. Allele frequency and consequently power of aim 2

Allele frequencies vary between CNV regions. Aim 1 is powered to study rare CNVs; in aim 2 we focus on CNVs with allele frequencies over 0.5%. Of note, genotyping error rates increase with increasing variant frequency, as the calculated reference is less likely to be exactly diploid. This diminishes dynamic range and results in overlapping Kernel distributions. The aCGH with the 1 to 1 comparison overcomes this problem by achieving higher genotyping accuracy specifically for the common variants. In the gene expression eQTL-CNV study we detected variant frequencies in the 11.4-63% range (last column added in Table 5). Aim 2 proposes to find variants that are too small to be visible on SNP arrays (superior resolution), have a measurable effect on gene expression (enrichment for true positives) and common enough to be important at the public health level. Power calculation based on simulations from the preliminary CNV eQTL study is available in section C2c4.

3. Microarray design needs more details. Not clear how many events will be detected.

Results of the aCGH design is added on page 9-10. We evaluated the microarray performance in silico and an example is added in Fig. 9. Extensive experience with Agilent aCGH suggests that the in silico experiment is highly reproducible with the catalogue probes. Baylor College of Medicine Genetic Diagnostic Labs developed a microarray for exon level coverage in known disease causing genes associated with severe pediatric phenotypes. These microarrays detect 30-40 events per sample (personal communications). The array proposed here has 4x number of probes distributed throughout the genome targeting regulatory regions and exons. The resolution is high to detect small events and also detects larger events as well. The number of inferred CNVs per sample is estimated to be in the hundreds. The main analysis method proposed is to use numeric segmented data in the analysis; the Affymetrix 6.0 data generated 80,514 segmentation covariates.

4. Expertise for RNA seq

We created a consortium with Roswell Park Cancer Institute (Bioinformatics Department). Dr. Jianmin Wang has joined the team; he has extensive experience in RNA seq analysis methods. Furthermore, we performed single stranded RNA seq on four post mortem brain samples (RIN 5.7-7.5) to assess feasibility. Quality control measures are added in Fig 10.

5. Whether there is sufficient time to complete the iPSC experiments during the proposed period.

It appears controversial even between reviewers. Clearly we will not be able to follow up with functional studies on all the detected associations. We proposed to follow up on the top 1-2 candidates within this period which we may not be able to complete within 5 years as cell line development can take up to 6-9 months after the biopsy is obtained. We have been able to secure a small grant to perform preliminary iPSC experiments and 4 lab members completed a 2-day iPSC practical course. We revised aim 3 as future directions.

Clarifications:

No new genotyping data is being generated. High quality CNV genotyping data is generated in aim 2. GWAS platforms have high genotyping error rates for CNVs, thus aim 2 using the gold standard aCGH with an expression guided design (in contrast to haplotype block driven design of SNP arrays) interrogates the genome in a novel way for CNV association.

Aspects of variability of brain tissue is not well addressed. In the gene expression analysis we will add post mortem interval (PMI), age, sex and RIN as covariates (page 11). As the experimental design uses CNV genotype and gene expression eQTL phenotype from the same individual, the contribution of non-genetic factors such as drug exposure, comorbidities, agonal state are less likely to cause a systemic bias.

The investigators keep referring to the significance of their dataset as "sporadic AD". It is unclear that these cases are actually sporadic (no family history) or merely isolated cases on ascertainment. The cases are isolated on ascertainment; we corrected it in the text.

A. SPECIFIC AIMS

Alzheimer's disease (AD) is a progressive neurodegenerative disease with an estimated heritability of 60-80%. Large scale genome-wide association studies (GWAS) using high frequency SNP variants identified 10 loci which do not account for the measured heritability. To find the missing heritability systematic assessment of all mutational mechanisms needs to be performed. Between the powerful SNP-GWAS studies and the planned Whole Genome Sequencing projects the contribution of copy number variation (CNV) to the genetic architecture of AD has not been studied fully. Although a limited number of CNV-GWAS studies have been reported based on SNP array data, these studies have not addressed the contribution of CNVs to the heritability of AD in at least two aspects: i) there is a need for studies with adequate sample size, e.g. that are powered to detect associations from CNVs with allele frequencies in the rare-intermediate range; and ii) there is a need for studies with adequate resolution. We propose to narrow the gap by performing an optimally powered study on existing datasets while also improving the resolution using innovative strategies (Aim 1) and optimizing resolution with a custom designed high density array comparative genome hybridization (aCGH) while also optimizing power by using gene-expression as quantitative trait locus (eQTL) (Aim 2).

Specific Aim 1. **Refine the contribution of CNVs to the genetic architecture of AD by performing a CNV GWAS using the Alzheimer's Disease Genetic Consortium dataset.**

Hypothesis: By applying alternative methodologies to existing SNP GWAS datasets (ADGC) the resolution of CNV detection can be improved and the contribution of CNVs to the genetic architecture of AD refined.

By optimizing normalization methods and logR ratio calculations, performing segmentation only to reduce the dataset where events may occur, performing the test of association on the numeric segmented data, the resolution and power is markedly improved. These provide a sample size of 7482 AD and 5333 normal controls, the largest sample size for CNV analysis so far. Discovery I will use the Affymetrix 6.0 datasets for best resolution and dynamic range among the SNP arrays; Discovery II will use ADC1-3 datasets captured on Illummina 660 arrays for still acceptable resolution with higher power; and the remaining datasets for locus specific replication. The segmentation without calls will allow the detection of smaller events, limited by the genomic coverage of the SNP array used.

Specific Aim 2. **Elucidate the contribution of small CNVs to the genetic architecture of AD by performing a CNV-eQTL association study on 200 AD and 200 normal control temporal lobes using custom Agilent array and Illumina RNA-seq**.

Hypothesis: By applying the gold standard method (aCGH) for CNV detection with a rationally-designed coverage and using eQTL to augment power the contribution of small CNVs to the genetic architecture of AD can be elucidated.

Existing datasets do not have the resolution to detect CNVs in the 150 bp-50 kb range and WGS will have limitations to assemble events in this size range due to the short reads, creating a gap in the size of CNVs assessed. In a genetically heterogeneous disorder, utilizing eQTL and CNVs as a genetic marker map within the same individuals in the context of case control status is a robust method to elucidate meaningful associations.

A) SIGNIFICANCE

A1) Importance: AD is the most common form of dementia and leads to unrelenting cognitive decline [1]. With increased longevity the prevalence of AD in the elderly represents a major public health problem. The heritability of AD is estimated at 60-80 % [2] forecasting potential of using genetic biomarkers for risk stratification in the future. The main risk factor for late-onset AD is the APOE4 allele with a population attributable fraction of 0.2-0.3 [3]. Several large scale genome-wide association studies (GWAS) using high frequency variants identified 9 additional loci with a combined population attributable fraction of 0.31 [3]. To find the missing heritability systematic assessment of all mutational mechanisms needs to be performed.

A2) Critical barriers: Between the powerful SNP-GWAS studies and the planned Whole Genome Sequencing projects we have not adequately assessed the contribution of copy number variation (CNV) to the genetic architecture of AD. CNV studies leveraging the SNP arrays used in traditional GWAS face multiple challenges, including variable coverage per platform, batch effects, and limited resolution due to inferior dynamic range [4]. To overcome these difficulties, CNV analyses of SNP arrays in AD applied very similar workflows, concentrating on high stringency calls [5-10]. There is a need for optimally powered studies using existing GWAS data, for studies that fully explore the existing GWAS data, and studies applying high density array comparative genome hybridization (aCGH) to evaluate small CNVs.

A3) Improvement of scientific knowledge: AD is a complex disease with insidious onset marked by significant neuronal loss by the time symptoms are observed and is superimposed on normal aging, thus AD has been very challenging to diagnose early and treat. Understanding the genetics of AD can contribute to the development of the field in various ways. (i) As genetic biomarkers are mostly stable, they can serve as biomarkers for risk stratification and early diagnosis in order to facilitate early intervention and success of disease modifying therapy, and (ii) genetics can delineate subsets within AD, and suggest subset specific treatment options. The failure of the numerous clinical trials over the last two decade prompts us to rethink and to try alternative approaches. GWAS studies suggest that AD is genetically heterogeneous, which likely implicates heterogeneity in disease pathogenesis as well. Genetic risk/pathomechanism specific treatment could result in success.

B) INNOVATION

CNVs contribute to genetic variance and serve as an alternative marker map for association studies. This alternative marker map can identify novel hypotheses for disease mechanism with an unbiased approach. Furthermore, each aim applies innovative strategies.

Aim 1. The existing datasets will be analyzed with alternative strategies, including methods developed by our team [11-13]. The workflow is depicted in Fig 4.

Aim 2. A) We will focus on joint analysis of CNV and gene expression. Based on the hypothesis that in a genetically heterogeneous disorder such as AD, utilizing eQTL and CNVs as a genetic marker map within the same individuals in the context of case control status may increase the power to detect relevant loci, we developed a statistical method which incorporates the linear effect for CNV together with a shift term for case-control status[12]. **B)** The high resolution array design with enrichment for exons and regulatory sequences wisely utilizes the probe allowance and generates a high resolution screen for the functional areas of the genome. **C)** The transcriptome sequence data is a valuable resource for the scientific community.

C) APPROACH

The two related and overlapping aims are directed to elucidate or exclude the role of CNVs contributing to the genetic architecture of AD with marked improvement in power and resolution compared to current data and analysis approaches.

C1) Aim 1. Refine the contribution of CNVs to the genetic architecture of AD by performing a CNV GWAS using the Alzheimer's Disease Genetic Consortium dataset. *Hypothesis: By applying alternative methodologies to existing SNP GWAS datasets (ADGC) the resolution of CNV detection can be improved and the contribution of CNVs to the genetic architecture of AD refined.*

C1a) Background: Six CNV case control GWAS studies have been published [5-10] (Table 1) and the ADGC preliminary analysis was presented at the 2011 ICAD meeting.

Table 1. Published CNV GWAS studies; case-control

Study	Platform	Input DNA	AD	MCI	Control
GERAD	Illumina 610-quad	200 ng	3260	0	1290
ADNI	Illumina Human610-Quad	NA	288	183	184
Caribbean Hispanics	Illumina HumanHap 650Y	NA	559	0	554
Duke	Illumina Human Hap550K	NA	331 *	0	368
TGEN	Affymetrix 6.0	NA	1022	0	595
NCRAD	Illumina Human610-Quad	NA	711	0	171

The majority of the studies performed genotyping on the Illumina platform with a coverage in the 600k range, except the TGEN study which used the Affymetrix 6.0 array with 2 Million probes (Table 2). The analysis methods were strikingly similar. To comply with the high stringency inferred CNV principal shared in all the studies, CNVs were excluded from the analysis based on number of probes, size and overlap with CNV variant regions or segmental duplications, sometimes even based on frequency (Table 2). These studies addressed only the tip of the iceberg with high specificity but low sensitivity for CNV detection.

Table 2. Outline of the methodology applied in published CNV GWAS studies; case-control

Study	LogR calculation	Reference file	Segmentation algorithm	Model	CNV exlusion
GERAD	BeadStudio	not mentioned	PennCNV	Hidden Markov Model	< 20 probes, <100 kb, density <1/15kb, >50% overlap with segdup
ADNI	GenomeStudio	not mentioned	PennCNV	Hidden Markov Model	<10 probes, overlap with centromer and immunoglobulin regions
Caribbean Hispanics	BeadStudio	not mentioned	QuantiSNP, iPattern, PennCNV, CNVpartition	Multiple	< 5 probes, <100 kb, overlap with centromer and immunoglobulin regions, 50% overlap with segdup, >1% frequency
Duke	BeadStudio	not mentioned	PennCNV	Hidden Markov Model	10 SNPs, 50% overlap with previously published regions
TGEN	unknown	not mentioned	PennCNV	Hidden Markov Model	10 SNPs, 50% overlap with centromeric, telomeric and immunoglobulin regions
NCRAD	GenomeStudio	not mentioned	PennCNV	Hidden Markov Model	<10 probes, likelihood ratio <10, centromer, immunoglobulin

Our group used alternative strategies to perform CNV GWAS analyses [11-13]. We would like to discuss an example as proof of principle.

C1b) Preliminary data: Genome-wide scan for Copy Number Variation association with AAO of AD

We performed a study of CNV association with AD AAO designed to examine the role of low-frequency variants with intermediate penetrance in the genetic architecture of AD[13]. AAO serves as a quantitative endophenotype and the cases only study design eliminates misclassification bias in this common disease with age-dependent penetrance. The alternative analysis strategy applied here is to use segmentation only to reduce the dataset where events may occur, perform the test of association on the numeric segmented data and validate the CNV calls if a replicated association signal was detected. This approach detects association signals from smaller events that would have been discarded when performing the high confidence calls and overcomes the need to determine exact dosage, which is often problematic at common CNV loci as the reference may deviate from the diploid state.

Five CNV regions were detected with sizes ranging from 3.6-24.8 kb and allele frequencies of 0.001-0.006 suggesting that small events with low frequencies could contribute to the genetic architecture of AD (Table 3).

Table 3. Cox proportional hazard regression using the univariate segmentation covariates as predictor and AAO as outcome. For the CNVs that showed an association (FDR<0.05) Cox hazard regression was repeated with adding gender and APOE (count of APOE4 alleles as categorical values) as covariates.

Probe	Chr:Mb	Size bp	ProbChiSq	fdr_p	ProbChiSq covariates	Nearest Gene	Carriers AAO mean	Non-carriers AAO mean
SNP_A-8600234	2:140	6125	6.4626E-06	0.0412	9.043E-06	None	50	71.57
SNP_A-8329031	3:131.9	3571	1.3493E-06	0.0330	1.609E-06	CPNE4	48.5	71.63
CN_1082571	4:42.7	6309	6.5114E-08	0.0051	1.992E-07	ATP8A1	56.4	71.7
SNP_A-8584575	8:139.9	4580	8.0255E-06	0.0421	5.911E-06	COL22A1	50	71.57
SNP_A-8327917	9:11.9	24789	4.3441E-06	0.0412	1.838E-05	None	49	71.63

The intragenic *CPNE4* deletion has been documented in the database of Genomic Variants and was validated in this set (Fig. 1). Due to size and probe coverage in the region this CNV was discarded in all the published studies, as well as the other regions detected.

C1c) Rationale: We propose to address the question whether CNVs contribute to the genetic architecture of AD by performing analyses on existing datasets (ADGC) by alternative methodologies. This approach addresses the power issue: the sample size is estimated from the ADGC dataset [3] as 7482 cases and 5333 controls; the studies are assigned to discovery versus replication cohorts based on the genotyping platform. CNV studies piggy-backing on SNP arrays face multiple challenges [4]. We hypothesize, that optimizing methods could overcome some of the challenges. In view of the resources already used for these SNP GWAS studies, we have to fully explore the possibility to use it for CNV analysis. Challenges that we face and the proposed method(s) to overcome these challenges are the following:

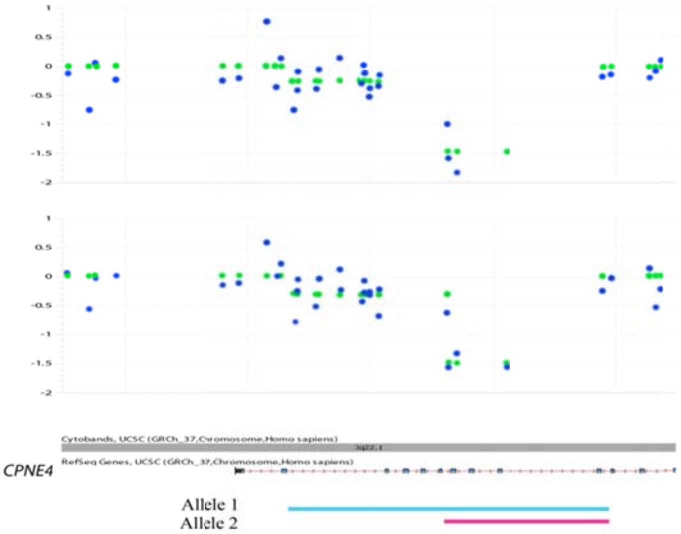


Figure 1. LogR data depicting the intragenic deletion in *CPNE4;* the upper panel shows a sample that is homozygous for the larger deletion (blue) and the lower panel depicts a compound heterozygote for the large (blue) and small (pink) deletion.

C1c1) Coverage: The various platforms have different coverage due to number of probes and principles of assay design [4]. We propose to use the highest density assay cohorts for discovery and the alternative assay cohorts for replication. This approach satisfies the independent cohort-independent assay principal for replication of an association study and maximizes coverage. We propose to use the higher density Affymetrix arrays for discovery and the Illumina arrays for replication. As the analysis workflow is the same for both the discovery and replication cohorts, the Illumina set will also serve as discovery for larger and less frequent CNVs.

Pitfalls and Alternatives: If the region is not covered on the Illumina arrays, we will not be able to use the Illumina cohort for replication. In these cases, we will develop a region and event specific high throughput assay (FISH, long-range PCR, MLPA assay, or qPCR) for the CNV and genotype the University at Buffalo and Baylor College of Medicine cohort (Table 4); subjects were not included in the microarrays.

C1c2) Batch effects: CNV analysis methods are sensitive to batch effects. The batch effects are related to amount of DNA used for the hybridization, the type of platform used and which laboratory performed the assay[4]. Often traditional normalization methods do not overcome the issue. However, in our experience the key is the logR ratio calculation [4]. As the SNP arrays compare a single genome to a reference file generated from multiple arrays, it is prudent that the reference file is generated within study. For example, for the TARC dataset if we use the HapMap controls reference file, a high number of arrays fail the MAPD QC measure. However, if we generate a reference file from the TARC controls, most arrays pass this QC criteria. We propose to perform logR ratio calculations within the sets resulting in a larger sample size due to less failed arrays and decreased confounders from batch effects both of which results in increased power.

Pitfalls and Alternatives: Applying different reference files could result in systemic changes within a study in regions where the CNV event is frequent. This can be resolved by comparing the reference files to each other. In addition, we will incorporate covariates into the analysis that will account for site/batch effects.

C1c3) Resolution: The CNV studies reported to date used high stringency calls [5-10]. This meant, based on the probe coverage and dynamic range of the various SNP arrays, that high-stringency calls were achieved for events that were over 100 kb in most studies. We propose segmentation to reduce the dataset while retaining the numeric data without calls [13]. The segmentation without calls allows the detection of smaller events, limited by the genomic coverage of the SNP array used. The segmentation reduces the dataset to areas of the genome where any CNV event may occur and discards the probes where there is no dosage variance. The reduction in comparisons decreases the multiple testing burden while maintaining the numeric data results in high sensitivity for CNV detection and increased power over the binned calls. The expectation is that all associations will be manually curated and replicated associations validated with orthogonal methods that are most appropriate for the region (FISH, long-range PCR, MLPA assay, or qPCR).

Pitfalls and Alternatives: The segmentation without calls may detect associations from noise which are eliminated through the manual curation process. False positive CNV calls if by chance are replicated are eliminated through the orthogonal validation step.

C1c4) Power calculation. The proposed discovery set has an 80% power to detect allelic association for an allele frequency/odds ratio combination of 0.005/2 or 0.01/1.6. (Harvard Genetic Power Calculator website)

C1d) Methods

C1d1) Data access. Deidentified data is obtained through collaboration. ***Intensity data access is available for the studies in table 4 (please find letters of collaboration attached).***

Table 4. ADGC datasets

Cohort	**Cases**	**Controls**	**Platform**	**dbGAP**	***Access***
TARCC	**375**	**198**	**Affymetrix 6.0**	**no**	***yes***
TGEN	**1022**	**595**	**Affymetrix 6.0**	**no**	***yes***
ADNI	**268**	**173**	**Illumina 610**	**no**	***Yes (letter attached)***
NIA-LOAD	**1811**	**1575**	**Illumina 610**	**yes**	***Yes***
ROSMAP	**296**	**776**	**Illumina 1M**	**yes**	***Yes (NIAGADS, letter attached)***
MIRAGE	**509**	**753**	**Illumina 610**	**no**	***Yes (NIAGADS, letter attached)***
ADC (1, 2, & 3)	**3201**	**1263**	**Illumina 660**	**no**	***Yes (NIAGADS, letter attached)***
Discovery I. Affy 6.0	***1397***	***793***	***High resolution, lower power***		
Discovery II. Illumina 660	***3201***	***1261***	***Lower resolution, higher power***		
Replication Illumina 610	***2884***	***3277***	***For CNVs with adequate coverage***		
Orthogonal locus specific assay	***800***	***800***	***For CNVs not covered by Illumina 610***		***UB, BCM; existing, non-overlapping with GWAS***

C1d2) Quality control*.* Experimental quality control. Affymetrix array: contrast QC and median absolute pair-wise differences; Illumina array: mean, median and s.d. of logR ratio and B allele frequency. Arrays with CNV calls more than two SD from the mean are eliminated.

Data quality control. Samples are excluded for low SNP call rate (<97%), gender mismatch (by X chromosome logR ratio), and related or duplicate samples (Pi >0.95) by determining IBD with PLINK software using the genotype data. From family based studies the proband will be selected.

C1d3) Population substructure/admixture by the SNP dataset. Eigensoft package is used to calculate PC from the SNP data and assign race/ethnicity cluster. Primary analysis will focus on Caucasian subjects; in addition the PCs are used as covariates in the statistical model.

C1d4) logR ratio calculation. As a first pass we generate the reference file from all controls in the given dataset. Second pass the reference for the complete analysis is generated from the top 100 DLRS control samples to optimize the elimination of noise. Normalization of logR data is performed by cRMAv2 (Bioconductor). The logR ratio data is subjected to numeric principal component analysis (GoldenHelix) and corrected for the number of PCs which yields a QQ plot devoid of inflation.

Progress: We performed a pilot study of normalization and segmentation in 4 different datasets. We used 20 CEL files of each dataset and calculated the logR using the reference virtual genome derived from the within study 20 CELfiles. We performed normalization using the cRMAv2 open resource package available at Bioconductor. The normalization includes calibration for allele probe pair crosstalk, nucleotide position effect and normalization for PCR fragment-length effect (Figure 2). The distribution of logR data is comparable and can be subjected to numeric analysis without binning.

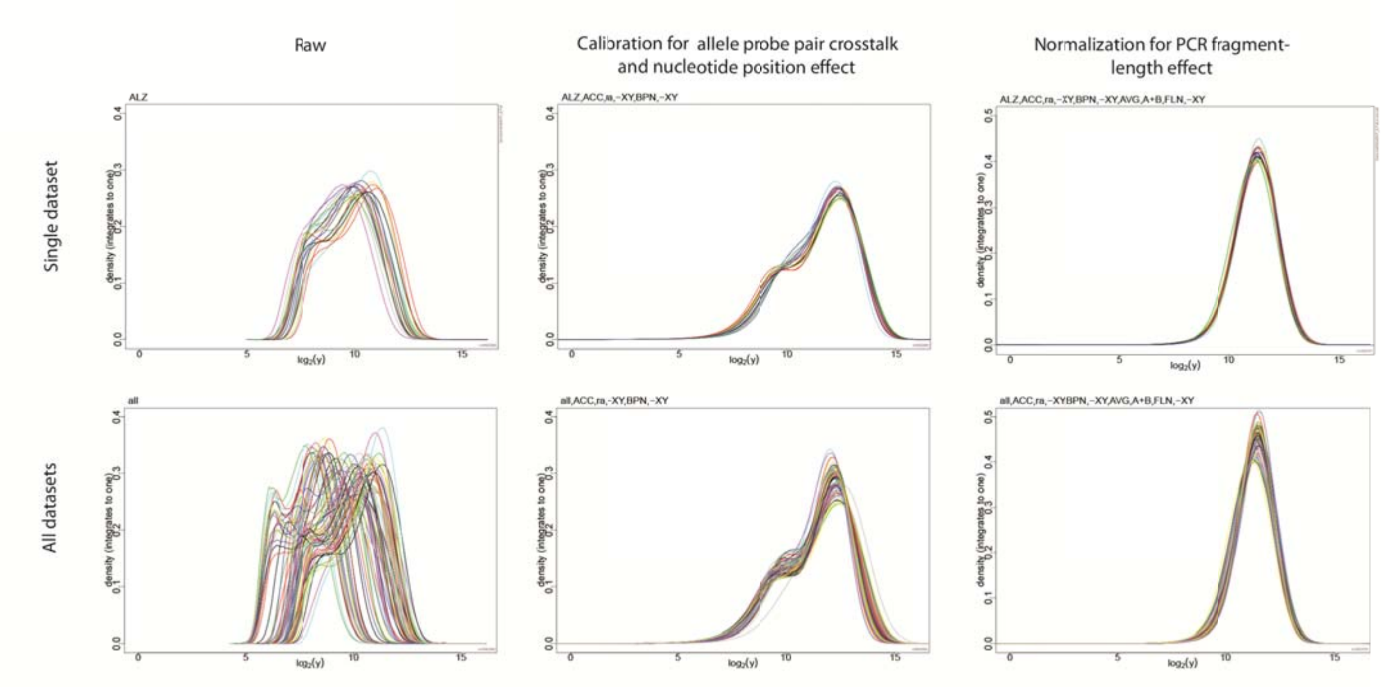


Figure 2. Normalization of Affymetrix 6.0 data from 4 different datasets. 20 CEL files from each dataset were used for logR calculation against a within dataset reference. The upper panels show a single dataset and the lower panel all 4 datasets together. The raw data (left), normalized data after calibration for allele probe crosstalk and nucleotide position effect (middle) and subsequent normalization for PCR fragment length effect (right) is visualized.

C1d5) Case control association analysis. The case control association is performed using logistic generalized linear model for the called CNVs and correlation-trend test for the numeric data.

C1d5a) Numerical array data: We created the analysis program in R to perform case control association with the numerical CNV array data. This approach searches for genomically contiguous regions where CN state has an effect on case control status. To enhance the analysis we take a "thin and bin" approach.

Thinning and Binning. Every other oligo is sampled to divide the data in half. In each half, *K* genomically adjacent oligos are binned and case-control association is performed on the mean CNV state within each thinned bin. FDR values for each thin bin p value is calculated, and the q-values for the CNV's coefficient from lowest (near 0) to highest (near 1) in each half is ranked. K=2 and K=100 is tested empirically. The K at which maximum concordance is attained with FDR q values less than 0.05 in each data half is selected. The direction of effect (sign of the beta coefficient) is verified to be concordant.

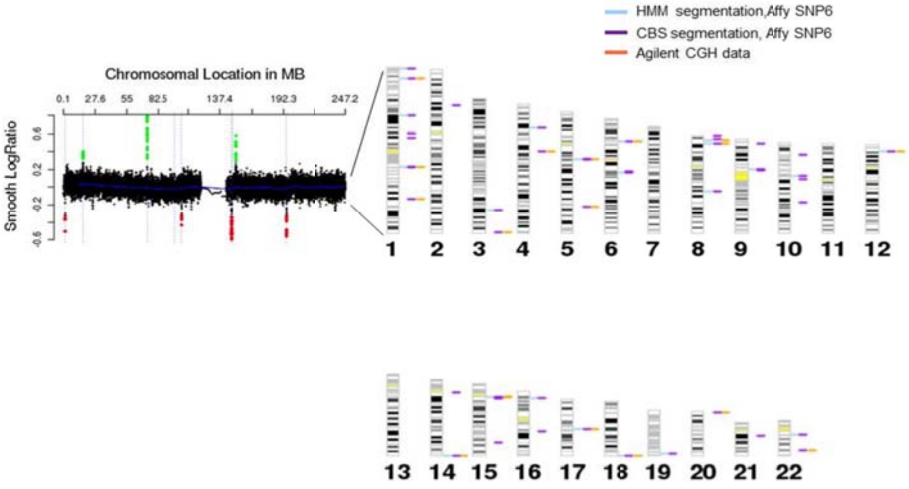
Effects of moderate size. The case control association is performed on the entire dataset removing the thinning but retaining data aggregation into *K* oligo bins. FDR q values are calculated.

C1d5b) Segmented numeric data: Numeric PC corrected data is segmented to identify probes where a CNV is detected in any of the samples in the set. The segmentation results in a reduced dataset while maintaining the advantages of the numeric data without binned CNVs. FDR q values are calculated.

C1d5c) Inferred CNVs: Two independent CNV calling algorithms are used. At this point in time the best performing algorithms for Affymetrix 6.0 data are Birdsuite and Ipattern[4]. We will perform the case control association on the CNV dosage. FDR q-values are calculated.

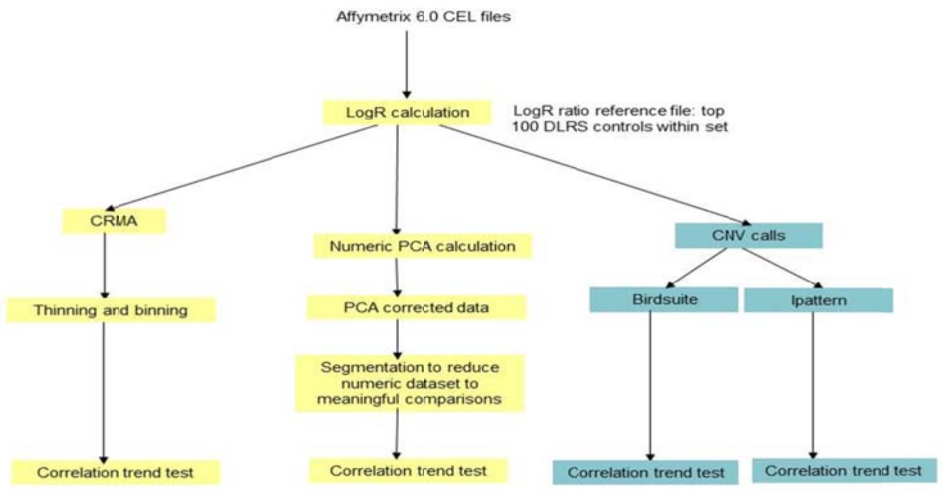
Preliminary data in our lab on CNV calling algorithms applied to Affymetrix 6.0 data and validated by the gold standard method of aCGH suggest that the sensitivity of the various algorithms to detect CNVs depends on the properties of the algorithm and the CNV, recently also reported in the literature[4] (Figure 3).

Figure 3. Segmentation of the Affymetrix 6.0 logR data with a Hidden Markov Model and Circular Binary Segmentation algorithm. CNVs were validated by array comparative genome hybridization (aCGH). The segmentation algorithms have different sensitivity to CNV events thus complement each other.



As the first pass CNV-GWAS analysis with high-stringency calls have been completed by other groups, we propose to enter all CNVs inferred by any of the algorithms into the analysis. The false positive associations will be eliminated by manual curation and orthogonal laboratory methods. Multiple testing correction is performed by the FDR approach. For the top candidates the logistic generalized linear model is repeated with sex and number of APOE4 alelles as covariates.

Figure 4. Data analysis plan



C1d6) Replication. The same workflow is applied to the Illumina dataset. During the interpretation, the coverage of the CNVs identified in the discovery set is assessed. For CNVs with suboptimal coverage calls are generated based on the Kernel densities, with the expectation that the allele frequencies are comparable to that detected in the discovery set [12]. For CNVs identified in the discovery set that have no coverage on the Illumina arrays, we design locus specific high throughput orthogonal assays (long-range PCR, MLPA assay, or qPCR) and perform the replication on the UB/BCM samples.

C2) Aim 2. Elucidate the contribution of small CNVs to the genetic architecture of AD by performing a CNV-eQTL association study on 200 AD and 200 normal control temporal lobes using custom Agilent array and Illumina RNA-seq. *Hypothesis: By applying the gold standard method (aCGH) for CNV detection with a rationally-designed coverage and using eQTL to augment power the contribution of small CNVs to the genetic architecture of AD can be elucidated.*

C2a) Background: Integrated Copy Number and Gene Expression Analysis
Genetic variation, both single nucleotide variations (SNV) and CNV, contribute to changes in gene expression. In some cases these variations are meaningfully correlated with disease states [14]. As GWAS studies are performed with increasing sample sizes [3, 15-17] it is becoming clear that in disorders with marked genetic heterogeneity where the marker specific risk is low in case-control sets, it is difficult to identify the true positives

from the false positives and to replicate the results [18,19]. In addition, case-control design in AD suffers from additional confounders, such as misclassification bias due to age-dependent penetrance. The approach of using gene-expression data from pathologically ascertained cases and controls reduces the misclassification bias and gene expression serves as a refined phenotype in a heterogeneous disorder, both increasing the power to detect association signals. We hypothesized that in a genetically heterogeneous disorder such as AD, utilizing eQTL and CNVs as a genetic marker map within the same individuals may increase the power to detect relevant loci, and by incorporating case control status, these loci will be candidates for case control association studies [12].

C2b) Preliminary results: In a pilot study of 35 post mortem temporal lobe samples, transcribed sequences were identified with evidence of differential expression between AD and controls. Five genes were both differentially expressed between cases and controls and had more than 50% of the variance attributable to a cis-CNV state (Table 5). Three of the six probes are adjacent to CNVR 1123.1 and fall into a linked region to AD on chromosome 2.

Table 5. Regression results (F-statistic p-values) for the component of case-control variance explained by CNV in the ANCOVA.

Probe_Id	ILMN_Gene	Chr	pADCNV	FracColinear	nVariants	Frequency
ILMN_1698680	ARL17P1	17:41.7	0.0004	0.90	22	62.8%
ILMN_2049364	FAM119A	2:208.1	0.0054	0.77	9	25.7%
ILMN_2334242	CREB1	2:208.1	0.0066	0.70	9	25.7%
ILMN_2334243	CREB1	2:208.1	0.0070	0.68	9	25.7%
ILMN_1696065	SDF4	1:1.1	0.0074	0.64	4	11.4%
ILMN_2155719	NBPF10	1:16.7	0.0098	0.53	31	88%

Two of these probes are replicates of *CREB1* and the third probe is *FAM119A*, a gene adjacent to *CREB1*.

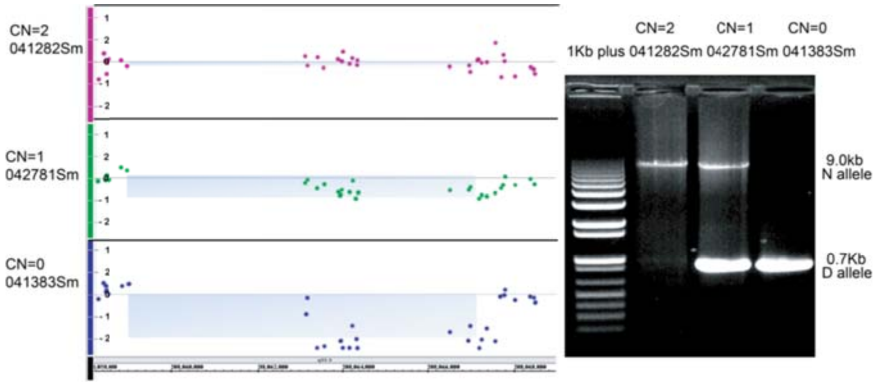


Figure 5. Validation of array data with long-range PCR.

The sign and size of the regression coefficients and p-values are similar suggesting that the association is observed in triplicate. Genotyping was confirmed by long range PCR (Fig. 5).

Subsequently, the associated CNVs were followed up in a case control association study (1230 AD subjects and 936 normal controls) and the deletion increases the risk of AD with an odds ratio of 1.23 (1.02-1.49). An independent dataset confirmed *CREB1* overexpression in AD (GSE15222). SNP association and linkage disequilibrium analysis of the TARCC dataset revealed that none of the SNPs tag the deletion (r^2 max 0.39).

Sequencing of the breakpoints of 5 homozygous deletion samples suggests a replication dependent mechanism [20] as the inserted sequence was identified 400 kb downstream from the deletion making recurrent events less likely (Fig. 6). The linkage disequilibrium map and the identical breakpoints suggest that the deletion occurred on the ancestral haplotype. This haplotype structure likely increased the statistical power and allowed the detection of the association signal and explains why GWAS studies using single SNPs may have not been able to detect this association signal.

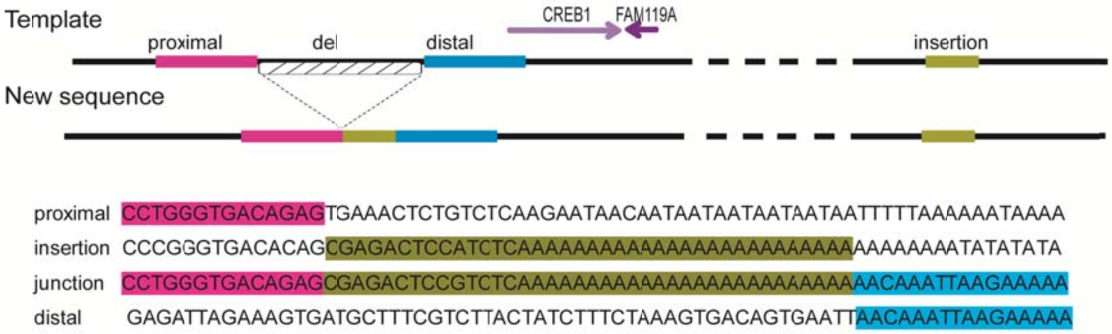

Template
proximal
del
distal
CREB1
FAM119A
insertion
New sequence
proximal CCTGGGTGACAGAGTGAAACTCTGTCTCAAGAATAACAATAATAATAATAATAATTTTTAAAAAATAAAA
insertion CCCGGGTGACACAGCGAGACTCCATCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATATATATA
junction CCTGGGTGACAGAGCGAGACTCCGTCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACAAATTAAGAAAAA
distal GAGATTAGAAAGTGATGCTTTCGTCTTACTATCTTTCTAAAGTGACAGTGAATTAACAAATTAAGAAAAA

Figure 6. Sequence of the breakpoint and inserted sequence; the insertion (gold) suggests a replicative mechanism for the deletion.

C2c) Rationale: In order to address the second unexplored question, whether small CNVs contribute to the genetic architecture of AD, we propose an aCGH gene-expression eQTL study.

C2c1) Why array CGH? The state of the art method to detect gene dosage variation is comparative array hybridization[4]. It has superior dynamic range due to the principal of comparing a test genome to a control genome, a 1 to 1 comparison[4]. Between SNP array based GWAS studies and large-scale WGS projects we still have not fully assessed CNVs in the 150 bp-50 kb range. Whole genome sequencing (WGS) will likely have limitations to assemble events in this size range due to the short reads. With the ability to custom design, enrichment for exons and regulatory sequences is achieved optimizing the coverage for the eQTL approach.

C2c2) Why eQTL? AD is a genetically heterogeneous disorder, which likely implicates heterogeneous pathomechanism as well. Gene expression improves the power in multiple ways: (i) the misclassification bias is practically eliminated, (ii) the gene expression endophenotype allows for disease heterogeneity in the analysis and (iii) the numeric data improves the statistical power in contrast to case control analysis.

Pitfalls and Alternatives: The utility of post mortem brain tissue for RNA studies has limitations. However, RNA is unusually protected in the brain post mortem, and even 24 hours after death good quality RNA can be isolated. Most RNA studies use covariates in the analyses to account for degradation, including post mortem interval. In the pilot study we maximized stringency by determining a threshold for RNA integrity number (RIN) to be included. We were able to obtain adequate quality of RNA (RIN > 4) in 70 percent of the cases. The detected association in the pilot study involved 3 adjacent RNA probes in 2 adjacent genes with very similar effect sizes and slopes representing a triplicate experiment, which is very reassuring.

C2c3) Why RNA sequencing? With the advent of next generation sequencing being applied to the transcriptome, a quantitatively and qualitatively superior level of information can be captured [21]. With the cost rapidly decreasing, such experiments became feasible. RNAseq provides superior dynamic range enabling the capture of five logs of expression level differences, thus abundant and rare transcripts can be detected in the same experiment. The sequence data can distinguish between splice variants and allelic variants, and provides information on non-coding RNA [21].

Pitfalls and Alternatives: The post mortem tissue may confound mainly the splice variant information due to degradation and can lead to false positives; however has been successfully studied in AD [21]. The degradation is expected to be random, thus present in the cases and the controls as well. In the statistical analysis we can correct for the degradation by using the RIN as a covariate in the model. Degradation could also affect quantification; however this can be controlled for by comparing the 3' and 5' exons.

C2c4) Power Calculation. We simulated data where a CNV is associated with a 1.5 fold increase in cis-expression of a gene that is otherwise 0.25 fold increased in AD versus controls. For a sample size of 200 cases and 200 controls the power exceeds 80% for allele frequencies of 0.05 and rises to over 95% when the CNV frequency is 0.08 or greater for situations where the CNV variant is approximately twice as common in cases as in controls*. **CNV frequencies detected in the preliminary studies varied between 0.001-0.88. Specific aim 2 proposes to find variants that are too small to be visible on SNP arrays (superior resolution), have a measurable effect on gene expression (enrichment for true positives) and common enough to be important at the public health level (aCGH offers superior dynamic range for common CNVs).***

C2d) Methods:

C2d1) Custom tiling aCGH.

Progress: The custom tiling oligo CGH array was designed and in silico tested. Genomic coordinates for all exons were downloaded from UCSC table format and resulted in 780 k exons. Genomic coordinates for brain expressed gene were obtained from Allan brain atlas; total of 17,995 genes. We

intersected the genomic regions of all exons with genomic regions of brain expressed genes; resulted in 180k exons. All regulatory regions in all cell lines form the ENCODE project was downloaded (wgEncodeRegTfbsClustered.bed.gz). 2.6 million regions were detected. These chromosomal regions were merged for non-overlapping regions which resulted in 600k genomic regions. Finally we concatenated the brain expressed exons with the non-overlapping regulatory regions, which resulted in 800k genomic regions. These regions were submitted to earray (Agilent) to identify catalogue probes. The search resulted in 970k probes.

To further ensure that brain regulatory regions are adequately covered we examined two LOAD and two healthy control samples for H3K27ac by ChIP-seq. H3K27ac is an indicative mark of active chromatin and is found at active promoters and enhancers. ChIP libraries were generated using ThruPLEX NGS library kit from Rubicon Genomics. The ChIP-seq results for H3K27ac are consistent with published results from the ENCODE project (Figure 7). The characteristic peaks of enrichment flanking the TSS are seen for all H3K27ac ChIP-seq experiments and are comparable to ENCODE. We identified 13,909 regulatory intervals. 80% of these regions overlapped with the ENCODE regulatory regions; the unique regions are added to the microarray design. Adding the backbone completed the 1million probe array design. Coverage of the microarray is depicted in Figure 8.

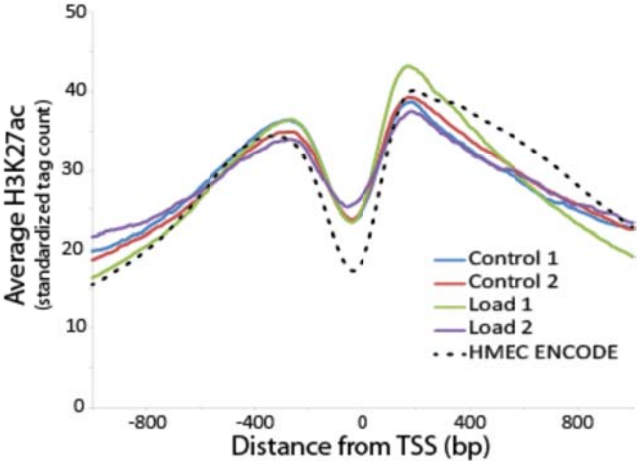


Figure 7. H3K27ac is enriched in temporal lobe samples at transcriptional start sites (TSS). The composite plots of H3K27ac at 27,259 gencode TSS was generated by extracting and extending aligned sequence tags. Plot was generated with our ArchTEx algorithm. Datasets were standardized to 50 million tags per library.

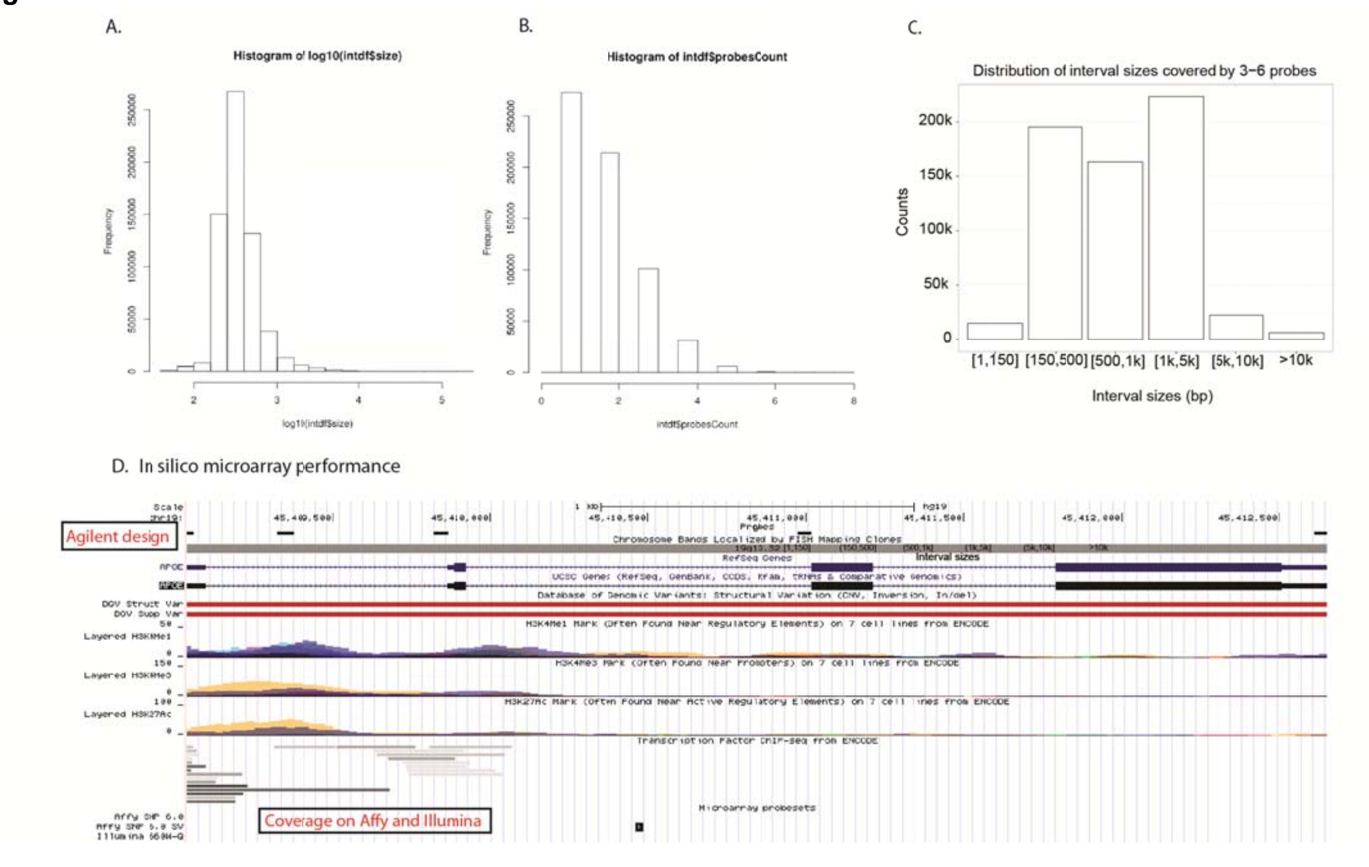


Figure 8. aCGH coverage and in silico performance. The size distribution (A, logscale) and the number of probes per interval distribution (B, logscale) of the 800k genomic regions are depicted . The distribution of interval sizes(in bp) covered by 3-6 probes in C reflects size resolution as the reproducibility of the CNV events covered by 3-6 probes by orthogonal assays is 70-95%. In silico performance of the designed microarray was evaluated for over 100 regions. One example is depicted in D. The APOE gene is covered by a single probe on the Affymetrix 6.0 array, while there are no probes on the Illumina 660 Quad.

The microarray presented here has 5 probes covering APOE, 3 in exons and 3 in the 5' end where regulatory regions are predicted. CNVs have been reported in the Database of Genomic Variants; GWAS platforms did not assay this region at all. Of note, the targeted high resolution coverage used one half of the probes as compared to the highest density Affymetrix 6.0 array. The resolution difference contrasts the array design principles: gene and regulatiory regions in the proposed microarray in contrast with the haplotype block driven design of the SNP arrays.

C2d2) Subject Cohorts. Human frozen temporal lobe tissue is obtained (Baylor College of Medicine, NIA funded Alzheimer Center brain banks [3] and the Miami and New York Brain Banks). 200 samples are available in the lab. Neuropathological diagnosis is assigned by board-certified neuropathologists. AD or control neuropathology is confirmed by plaque and tangle assessment and Braak staging. The brain cohort is exempt from IRB approval as the specimens are deidentified. Tissue request of 400 samples has been approved by the Miami Brain Bank; letter attached.

C2d3) Isolation of DNA and RNA from Brain Tissue, Expression Profiling and Genotyping. DNA and RNA is prepared by standard procedures (QIAamp DNA mini kit (Qiagen) and Trizol (Invitrogen) and RNeasy mini kit (Qiagen), respectively). DNA and RNA QC criteria for proceeding to array experiments are 260/280nm 1.6-1.8 for DNA and 1.9-2.0 for RNA, 260/230nm >2.0 for DNA and > 1.5 for RNA by Nanodrop 1000. RNA quality is further assessed by calculating RNA integrity number (RIN) with Agilent 2100. Sample is entered to the expression array experiment if RIN>4. Approximately 70% of samples achieve this RNA quality. Comperative genome hybridization is performed by standard protocol (Agilent). Paired-end RNAseq will be performed on the Illumina RNA-seq platforms by the core facility at the University at Buffalo. APOE genotyping is performed with real-time PCR using custom TaqMan probes (Applied Biosystems, Inc).

C2d4) Analysis of Array Data

C2d4a) aCGH Data. Normalized numeric array data, segmented numeric data and inferred CNV calls are used in the analysis (described in Aim 1). Normalization of the numeric array data is achieved by the quantile normalization method developed by our group. Segmentation (GoldenHelix) reduces the dataset to meaningful comparisons where any CNV event occurs but maintains the numeric data without binning into CN state. This approach maintains the power gain from the numeric data and reduces the multiple testing burden. CNV calls are based on the Agilent calling algorithm. For quality control the derivative log ratio spreads (dlrs) are calculated (<0.3).

C2d4b) Expression Data. Raw RNA-seq data is demultiplexed and aligned against the human genome (hg19) using the TopHat alignment tool [22]. ***The resulting bam file is analyzed by the cufflinks [23] program to estimate transcript and gene level abundance measured by FPKM (Fragments per Kilobase per Million Mapped Fragments) values. Cufflink program can also predict alternative isoforms by isoform assembly and searching and quantification all isoforms by statistical analysis. Software development is diligently monitored and the best available method(s) is used. Numeric exon level data measured by FPKM is entered into the analysis of the following step.***

Progress: We performed single stranded RNA sequencing from post mortem brain tissue. Two AD and two normal control samples were sequenced. RIN ranged 5.6-7.5. RNA was prepared with the TruSeq RNA kit (Illumina Inc), from 1µg total RNA using standard procedures. The samples were multiplexed and 50-bp single strand sequenced (Ilumina HiSeq2500). Prior to pooling, each individual sample's amplified cDNA construct were visualized on a DNA-HS Bioanalyzer DNA chip (Agilent Technologies). Successful libraries were purified and pooled at equal molar before sequencing. Raw RNA-seq data were demultiplexed and aligned against the human genome (hg19) using the TopHat alignment tool [22]. QC measures including coverage, relative coverage of exons, introns and intergenic sequences, correlation between samples (reproducibility) and GC bias were calculated [24] (Fig 9.)

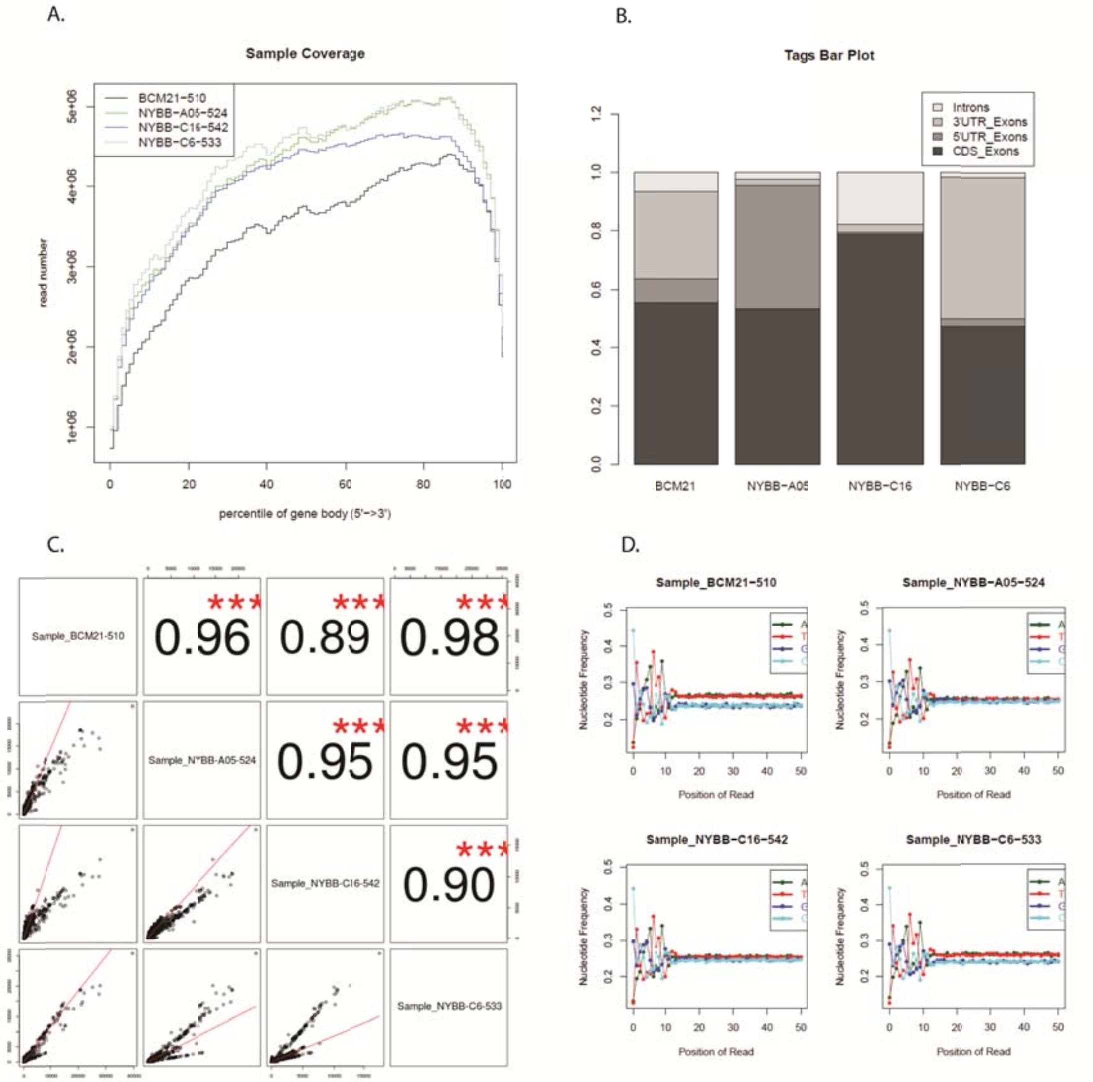


Figure 9. Single stranded RNA seq data QC measures are depicted for 4 post mortem temporal lobe samples, two AD and two normal controls. Coverage (A), relative coverage of exons, introns and intergenic regions (B), expression level performance as correlation between samples (C) and GC bias (D) reveal good quality sequencing results. The parameters will likely improve further by using RNase H method for library generation, post hoc GC normalization, deeper coverage and paired end seqencing.

C2d4c) Definition of CNV-Expression. UCSC browser track is utilized for mapping every transcript to the human genome Build 37; 500 Kb padding is added on each side of the transcript (1Mb total window) and any CNV calls overlapping the 1Mb windows are associated with that transcript. The 500 kb was selected based on previously published data [25] suggesting that within this distance the likelihood of an effect of a SNP on gene regulation is over 99%.

C2d4d) Statistical analysis incorporating expression, disease-state and CNV information. The following filters are applied to reduce the dataset to meaningful comparisons for the hypothesis: (i) differentially expressed between cases and controls (T-test, p<0.05; ***covariates include post mortem interval (PMI), age, sex and RIN)***, and (ii) CNV events in at least 3 of the individuals that are within 1 MB of the transcript (frequency>0.0075) to avoid spurious associations caused by single events at the extremes of the expression spectrum. Subsequently, we will perform ANCOVA analysis to identify situation where cis-CNVs are confounded with expression differences between cases and controls. To perform this analysis we have created the analysis program in R. The null hypothesis is that the difference in expression between cases and controls does not correlate with cis CNV state. ANCOVA is performed for each transcript where a linear effect for CNV together with a shift term for case-control status was estimated by least squares. The fraction of variation (partial R^2) attributable to CNV state was assessed by taking the ratio between the regression sum of squares for the CNV information to the regression sum of squares for the full model. The F-statistic and a corresponding p-value are determined by dividing the mean square error for CNV by the mean square error of the full ANCOVA.   The analysis will be performed with three CNV datasets: numeric, segmented numeric and binned CNV calls. For multiple testing correction, in addition to q-value computation we will perform 1000 permutations. Data in Aim 1 is used to test the hypothesis whether the identified CNV is a risk allele in case-control analysis. The top candidates are reanalyzed with age, sex and number of APOE4 alleles in the model.

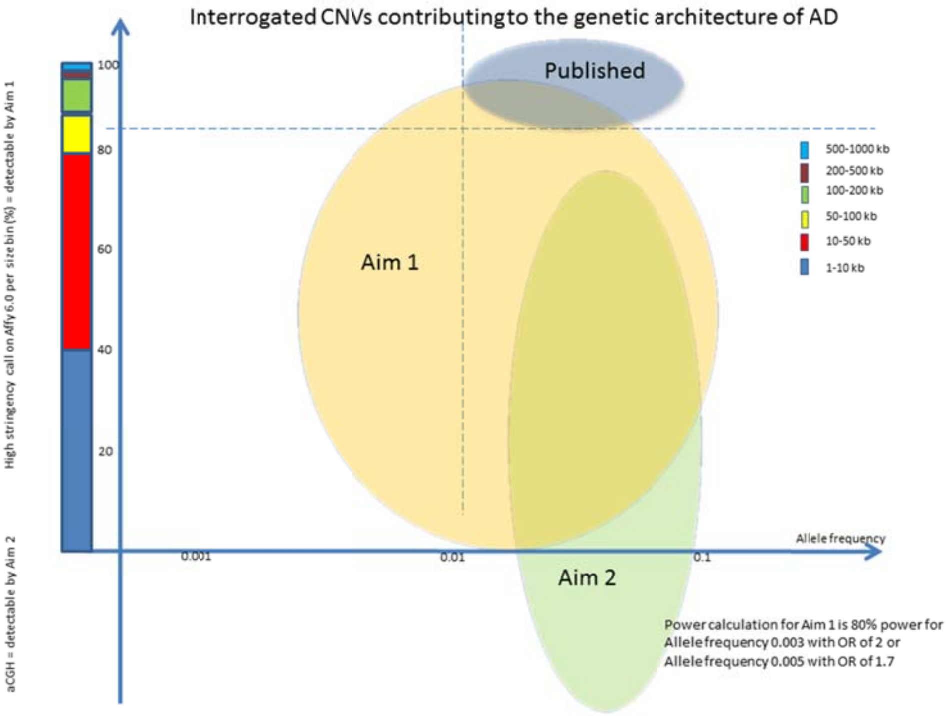
C2d5) Replication. Two tiers of the CNV eQTL replication studies will be performed depending on the size of the CNV, its coverage on Affymetrix 6.0 array and the abundance of the expression of the candidate gene. First tier: CNVs that are covered on the Affymetrix 6.0 microarray with at least 3-5 probes and are abundant for detection on microarray. We have access to the TGEN GWAS dataset on Affymetrix 6.0 of which a subset also has Illumina gene expression data in Gene Expression Omnibus (xx AD and XX normal controls). Canditate CNV-eQTL pairs will be subtracted from the corresponding dataset and the same analysis method will be applied. Multiple testing correction will be based on the number of candidate CNV-eQTL pairs identified. Second tier: As the proposed aCGH and the RNA seq experiments have high resolution, there will be candidate CNV eQTL pairs not covered on the existing datasets. For these CNV eQTL pairs we will develop orthogonal assays for CNV detection (TaqMan assay, long range PCR, MLPA) and qPCR for gene expression and splice variant analysis. The DNA and RNA from post mortem human brain is available through collaboration. Amanda Myers PhD will share her extensive gene expression data and will perform the CNV genotyping (Please see the letter of support attached).

FUTURE DIRECTIONS: **Functional validation of the top candidates using iPSC.** Functional validation of the effect of CNVs in the brain in AD is challenging at multiple levels. The size of CNVs often precludes cloning and the development of in vitro systems. Blood derived lymphoblastoid cell lines often do not correlate with gene regulation/expression in the brain due to tissue specific regulation. AD as a disease is challenging to model in animal systems. iPSC differentiated into neurons provides an in vitro assay system to study the effect of a given CNV on the pathomechanism of AD. After establishing readout and rescue, this in vitro system is also adapt for high throughput drug screen in a genotype and mechanism specific manner. ***Overlapping results from specific aims 1 and 2 are prioritized. Experimental design is driven by the CNV specific putative pathomechanism.***

C3) Overall Summary and Conclusion

The completion of this project will expand the depth of the inquiry whether CNVs contribute to the heritability of AD as depicted in Figure 10 and likely identify any important association from the public health standpoint.

Figure 10. Closing the gap in our knowledge on the role of CNVs in the genetic architecture of AD. The colored bar on the left represents the percentage of high stringency CNVs from each size range inferred from Affymetrix 6.0 data. Aim 1 investigates all these calls and the segmented numeric analysis expands it to smaller events as well limited by the coverage of Affy 6.0. The aCGH-eQTL in Aim 2 will have high sensitivity for small events down to the 150 bp range complementing Aim 1 and expanding to areas of the genome where Affy coverage is sparse.



Power calculation for Aim 1 is 80% power for Allele frequency 0.003 with OR of 2 or Allele frequency 0.005 with OR of 1.7

Timeline

This project would take 5 years to complete. Aims can be parallelized to some degree. In Aim 1 data analysis can start right away and continues in the first 2-3 years. For Aim 2 in the first year DNA and RNA are harvested followed by data capturing in years 2-4 and analysis in year 5.

PROTECTION OF HUMAN SUBJECTS

Human Subjects Involvement

This human subject research falls under exemption 4 as both the GWAS data and the post mortem temporal lobe tissue are deidentified.

Aim 1: Deidentified GWAS datasets are requested through the NIH data sharing policy. All studies used consent forms that asked the subject whether they would allow their data to be shared. University at Buffalo IRB is pending for the use of this deidentified information.

Aim 2: Post mortem human temporal lobe tissue is requested from NIA supported brain banks. These samples are deidentified. University at Buffalo IRB is pending for the use of this deidentified information.

1. Kukull, W.A., R. Higdon, J.D. Bowen, W.C. McCormick, L. Teri, G.D. Schellenberg, G. van Belle, L. Jolley, and E.B. Larson, *Dementia and Alzheimer disease incidence: a prospective cohort study.* Arch Neurol, 2002. **59**(11): p. 1737-46.
2. Gatz, M., N.L. Pedersen, S. Berg, B. Johansson, K. Johansson, J.A. Mortimer, S.F. Posner, M. Viitanen, B. Winblad, and A. Ahlbom, *Heritability for Alzheimer's disease: the study of dementia in Swedish twins.* J Gerontol A Biol Sci Med Sci, 1997. **52**(2): p. M117-25.
3. Naj, A.C., G. Jun, G.W. Beecham, L.S. Wang, B.N. Vardarajan, J. Buros, P.J. Gallins, J.D. Buxbaum, G.P. Jarvik, P.K. Crane, E.B. Larson, T.D. Bird, B.F. Boeve, N.R. Graff-Radford, P.L. De Jager, D. Evans, J.A. Schneider, M.M. Carrasquillo, N. Ertekin-Taner, S.G. Younkin, C. Cruchaga, J.S. Kauwe, P. Nowotny, P. Kramer, J. Hardy, M.J. Huentelman, A.J. Myers, M.M. Barmada, F.Y. Demirci, C.T. Baldwin, R.C. Green, E. Rogaeva, P. St George-Hyslop, S.E. Arnold, R. Barber, T. Beach, E.H. Bigio, J.D. Bowen, A. Boxer, J.R. Burke, N.J. Cairns, C.S. Carlson, R.M. Carney, S.L. Carroll, H.C. Chui, D.G. Clark, J. Corneveaux, C.W. Cotman, J.L. Cummings, C. DeCarli, S.T. DeKosky, R. Diaz-Arrastia, M. Dick, D.W. Dickson, W.G. Ellis, K.M. Faber, K.B. Fallon, M.R. Farlow, S. Ferris, M.P. Frosch, D.R. Galasko, M. Ganguli, M. Gearing, D.H. Geschwind, B. Ghetti, J.R. Gilbert, S. Gilman, B. Giordani, J.D. Glass, J.H. Growdon, R.L. Hamilton, L.E. Harrell, E. Head, L.S. Honig, C.M. Hulette, B.T. Hyman, G.A. Jicha, L.W. Jin, N. Johnson, J. Karlawish, A. Karydas, J.A. Kaye, R. Kim, E.H. Koo, N.W. Kowall, J.J. Lah, A.I. Levey, A.P. Lieberman, O.L. Lopez, W.J. Mack, D.C. Marson, F. Martiniuk, D.C. Mash, E. Masliah, W.C. McCormick, S.M. McCurry, A.N. McDavid, A.C. McKee, M. Mesulam, B.L. Miller, C.A. Miller, J.W. Miller, J.E. Parisi, D.P. Perl, E. Peskind, R.C. Petersen, W.W. Poon, J.F. Quinn, R.A. Rajbhandary, M. Raskind, B. Reisberg, J.M. Ringman, E.D. Roberson, R.N. Rosenberg, M. Sano, L.S. Schneider, W. Seeley, M.L. Shelanski, M.A. Slifer, C.D. Smith, J.A. Sonnen, S. Spina, R.A. Stern, R.E. Tanzi, J.Q. Trojanowski, J.C. Troncoso, V.M. Van Deerlin, H.V. Vinters, J.P. Vonsattel, S. Weintraub, K.A. Welsh-Bohmer, J. Williamson, R.L. Woltjer, L.B. Cantwell, B.A. Dombroski, D. Beekly, K.L. Lunetta, E.R. Martin, M.I. Kamboh, A.J. Saykin, E.M. Reiman, D.A. Bennett, J.C. Morris, T.J. Montine, A.M. Goate, D. Blacker, D.W. Tsuang, H. Hakonarson, W.A. Kukull, T.M. Foroud, J.L. Haines, R. Mayeux, M.A. Pericak-Vance, L.A. Farrer and G.D. Schellenberg, *Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease.* Nat Genet, 2011. **43**(5): p. 436-41.
4. Pinto, D., K. Darvishi, X. Shi, D. Rajan, D. Rigler, T. Fitzgerald, A.C. Lionel, B. Thiruvahindrapuram, J.R. Macdonald, R. Mills, A. Prasad, K. Noonan, S. Gribble, E. Prigmore, P.K. Donahoe, R.S. Smith, J.H. Park, M.E. Hurles, N.P. Carter, C. Lee, S.W. Scherer, and L. Feuk, *Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants.* Nat Biotechnol, 2011. **29**(6): p. 512-20.
5. Swaminathan, S., S. Kim, L. Shen, S.L. Risacher, T. Foroud, N. Pankratz, S.G. Potkin, M.J. Huentelman, D.W. Craig, M.W. Weiner, A.J. Saykin, and A. The Alzheimer's Disease Neuroimaging Initiative, *Genomic Copy Number Analysis in Alzheimer's Disease and Mild Cognitive Impairment: An ADNI Study.* Int J Alzheimers Dis, 2011. **2011**: p. 729478.
6. Heinzen, E.L., A.C. Need, K.M. Hayden, O. Chiba-Falek, A.D. Roses, W.J. Strittmatter, J.R. Burke, C.M. Hulette, K.A. Welsh-Bohmer, and D.B. Goldstein, *Genome-wide scan of copy number variation in late-onset Alzheimer's disease.* J Alzheimers Dis, 2010. **19**(1): p. 69-77.
7. Swaminathan, S., M.J. Huentelman, J.J. Corneveaux, A.J. Myers, K.M. Faber, T. Foroud, R. Mayeux, L. Shen, S. Kim, M. Turk, J. Hardy, E.M. Reiman, and A.J. Saykin, *Analysis of Copy Number Variation in Alzheimer's Disease in a Cohort of Clinically Characterized and Neuropathologically Verified Individuals.* PLoS One, 2012. **7**(12): p. e50640.
8. Ghani, M., D. Pinto, J.H. Lee, Y. Grinberg, C. Sato, D. Moreno, S.W. Scherer, R. Mayeux, P. St George-Hyslop, and E. Rogaeva, *Genome-wide survey of large rare copy number variants in Alzheimer's disease among Caribbean hispanics.* G3 (Bethesda), 2012. **2**(1): p. 71-8.
9. Chapman, J., E. Rees, D. Harold, D. Ivanov, A. Gerrish, R. Sims, P. Hollingworth, A. Stretton, P. Holmans, M.J. Owen, M.C. O'Donovan, J. Williams, and G. Kirov, *A genome-wide study shows a limited contribution of rare copy number variants to Alzheimer's disease risk.* Hum Mol Genet, 2012.
10. Swaminathan, S., L. Shen, S. Kim, M. Inlow, J.D. West, K.M. Faber, T. Foroud, R. Mayeux, and A.J. Saykin, *Analysis of Copy Number Variation in Alzheimer's Disease: the NIA-LOAD/NCRAD Family Study.* Curr Alzheimer Res, 2012.

11. Shaw, C.A., Y. Li, J. Wiszniewska, S. Chasse, S.N. Zaidi, W. Jin, B. Dawson, K. Wilhelmsen, J.R. Lupski, J.W. Belmont, R.S. Doody, and K. Szigeti, *Olfactory copy number association with age at onset of Alzheimer disease.* Neurology, 2011. **76**(15): p. 1302-9.
12. Li, Y., C.A. Shaw, I. Sheffer, N. Sule, S.Z. Powell, B. Dawson, S.N. Zaidi, K.L. Bucasas, J.R. Lupski, K.C. Wilhelmsen, R. Doody, and K. Szigeti, *Integrated copy number and gene expression analysis detects a CREB1 association with Alzheimer's disease.* Transl Psychiatry, 2012. **2**: p. e192.
13. Szigeti, K., D. Lal, Y. Li, R.S. Doody, K. Wilhelmsen, L. Yan, S. Liu, and C. Ma, *Genome-wide scan for copy number variation association with age at onset of Alzheimer's disease.* J Alzheimers Dis, 2013. **33**(2): p. 517-23.
14. Nicolae, D.L., E. Gamazon, W. Zhang, S. Duan, M.E. Dolan, and N.J. Cox, *Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS.* PLoS Genet, 2010. **6**(4): p. e1000888.
15. Harold, D., R. Abraham, P. Hollingworth, R. Sims, A. Gerrish, M.L. Hamshere, J.S. Pahwa, V. Moskvina, K. Dowzell, A. Williams, N. Jones, C. Thomas, A. Stretton, A.R. Morgan, S. Lovestone, J. Powell, P. Proitsi, M.K. Lupton, C. Brayne, D.C. Rubinsztein, M. Gill, B. Lawlor, A. Lynch, K. Morgan, K.S. Brown, P.A. Passmore, D. Craig, B. McGuinness, S. Todd, C. Holmes, D. Mann, A.D. Smith, S. Love, P.G. Kehoe, J. Hardy, S. Mead, N. Fox, M. Rossor, J. Collinge, W. Maier, F. Jessen, B. Schurmann, H. van den Bussche, I. Heuser, J. Kornhuber, J. Wiltfang, M. Dichgans, L. Frolich, H. Hampel, M. Hull, D. Rujescu, A.M. Goate, J.S. Kauwe, C. Cruchaga, P. Nowotny, J.C. Morris, K. Mayo, K. Sleegers, K. Bettens, S. Engelborghs, P.P. De Deyn, C. Van Broeckhoven, G. Livingston, N.J. Bass, H. Gurling, A. McQuillin, R. Gwilliam, P. Deloukas, A. Al-Chalabi, C.E. Shaw, M. Tsolaki, A.B. Singleton, R. Guerreiro, T.W. Muhleisen, M.M. Nothen, S. Moebus, K.H. Jockel, N. Klopp, H.E. Wichmann, M.M. Carrasquillo, V.S. Pankratz, S.G. Younkin, P.A. Holmans, M. O'Donovan, M.J. Owen, and J. Williams, *Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease.* Nat Genet, 2009. **41**(10): p. 1088-93.
16. Seshadri, S., A.L. Fitzpatrick, M.A. Ikram, A.L. DeStefano, V. Gudnason, M. Boada, J.C. Bis, A.V. Smith, M.M. Carassquillo, J.C. Lambert, D. Harold, E.M. Schrijvers, R. Ramirez-Lorca, S. Debette, W.T. Longstreth, Jr., A.C. Janssens, V.S. Pankratz, J.F. Dartigues, P. Hollingworth, T. Aspelund, I. Hernandez, A. Beiser, L.H. Kuller, P.J. Koudstaal, D.W. Dickson, C. Tzourio, R. Abraham, C. Antunez, Y. Du, J.I. Rotter, Y.S. Aulchenko, T.B. Harris, R.C. Petersen, C. Berr, M.J. Owen, J. Lopez-Arrieta, B.N. Varadarajan, J.T. Becker, F. Rivadeneira, M.A. Nalls, N.R. Graff-Radford, D. Campion, S. Auerbach, K. Rice, A. Hofman, P.V. Jonsson, H. Schmidt, M. Lathrop, T.H. Mosley, R. Au, B.M. Psaty, A.G. Uitterlinden, L.A. Farrer, T. Lumley, A. Ruiz, J. Williams, P. Amouyel, S.G. Younkin, P.A. Wolf, L.J. Launer, O.L. Lopez, C.M. van Duijn, and M.M. Breteler, *Genome-wide analysis of genetic loci associated with Alzheimer disease.* JAMA, 2010. **303**(18): p. 1832-40.
17. Lambert, J.C., S. Heath, G. Even, D. Campion, K. Sleegers, M. Hiltunen, O. Combarros, D. Zelenika, M.J. Bullido, B. Tavernier, L. Letenneur, K. Bettens, C. Berr, F. Pasquier, N. Fievet, P. Barberger-Gateau, S. Engelborghs, P. De Deyn, I. Mateo, A. Franck, S. Helisalmi, E. Porcellini, O. Hanon, M.M. de Pancorbo, C. Lendon, C. Dufouil, C. Jaillard, T. Leveillard, V. Alvarez, P. Bosco, M. Mancuso, F. Panza, B. Nacmias, P. Bossu, P. Piccardi, G. Annoni, D. Seripa, D. Galimberti, D. Hannequin, F. Licastro, H. Soininen, K. Ritchie, H. Blanche, J.F. Dartigues, C. Tzourio, I. Gut, C. Van Broeckhoven, A. Alperovitch, M. Lathrop, and P. Amouyel, *Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease.* Nat Genet, 2009. **41**(10): p. 1094-9.
18. Ku, C.S., E.Y. Loy, Y. Pawitan, and K.S. Chia, *The pursuit of genome-wide association studies: where are we now?* J Hum Genet, 2010. **55**(4): p. 195-206.
19. Florez, J.C., *Clinical review: the genetics of type 2 diabetes: a realistic appraisal in 2008.* J Clin Endocrinol Metab, 2008. **93**(12): p. 4633-42.
20. Hastings, P.J., G. Ira, and J.R. Lupski, *A microhomology-mediated break-induced replication model for the origin of human copy number variation.* PLoS Genet, 2009. **5**(1): p. e1000327.
21. Twine, N.A., K. Janitz, M.R. Wilkins, and M. Janitz, *Whole transcriptome sequencing reveals gene expression and splicing differences in brain regions affected by Alzheimer's disease.* PLoS One, 2011. **6**(1): p. e16266.
22. Trapnell, C., L. Pachter, and S.L. Salzberg, *TopHat: discovering splice junctions with RNA-Seq.* Bioinformatics, 2009. **25**(9): p. 1105-11.

23. Trapnell, C., B.A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M.J. van Baren, S.L. Salzberg, B.J. Wold, and L. Pachter, *Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation.* Nat Biotechnol, 2010. **28**(5): p. 511-5.

24. Wang, L., S. Wang, and W. Li, *RSeQC: quality control of RNA-seq experiments.* Bioinformatics, 2012. **28**(16): p. 2184-5.

25. Stranger, B.E., M.S. Forrest, M. Dunning, C.E. Ingle, C. Beazley, N. Thorne, R. Redon, C.P. Bird, A. de Grassi, C. Lee, C. Tyler-Smith, N. Carter, S.W. Scherer, S. Tavare, P. Deloukas, M.E. Hurles, and E.T. Dermitzakis, *Relative impact of nucleotide and copy number variation on gene expression phenotypes.* Science, 2007. **315**(5813): p. 848-53.

CONSORTIUM/CONTRACTUAL ARRANGEMENTS

This project will be conducted by Dr. Szigeti at the University at Buffalo. Dr. Wang at Roswell Park Cancer Institute will be in charge of overseeing the research efforts at Roswell Park Cancer Institute focusing on the statistical aspects of RNA seq data processing and interpretation of Specific aim 2. Dr. Yan will oversee the downstream statistical aspects of specific aim 2. Dr. Wang will function as the primary contact person there with respect to the general operations of the facility, whereas specific day-to-day operations will be managed via technical and research assistants. The appropriate programmatic and administrative personnel of Roswell Park Cancer Institute and the University at Buffalo are aware of the agency’s consortium agreement policy and are prepared to establish the necessary inter-organizational agreement consistent with that policy.

DEPARTMENT OF RADIOLOGY
AND IMAGING SCIENCES

INDIANA UNIVERSITY
School of Medicine

October 17, 2013

Kinga Szigeti, MD, PhD
Assistant Professor of Neurology
Director, Alzheimer's Disease and Memory Disorders Center
University at Buffalo, SUNY

Dear Kinga,

As Genetics Core Leader of the Alzheimer’s Disease Neuroimaging Initiative, I am writing to express my enthusiastic support for your proposed CNV analysis. You already have access to the intensity data files of the ADNI Illumina 610 GWAS study and have been working with the data.

Further research on copy number variation in Alzheimer’s disease is very important and there are few groups working intensively on this. I certainly encourage your pursuing these studies. I wish you the best of luck with your proposal and subsequent research.

Please do not hesitate to contact me should I be able to be of assistance in your research.

Sincerely,

Andrew J Saykin

Andrew J. Saykin, PsyD, ABPP/ABCN
Raymond C. Beeler Professor of Radiology and Imaging Sciences
Professor of Medical and Molecular Genetics
Director, Indiana Alzheimer Disease Center
Director, Indiana University Center for Neuroimaging
Editor-in-Chief, *Brain Imaging and Behavior* (www.springer.com/journal/11682)

UHealth
UNIVERSITY OF MIAMI HEALTH SYSTEM
Neurology

UNIVERSITY OF MIAMI
MILLER SCHOOL
of MEDICINE

QUOTE

VENDOR: Brain Endowment Bank
Dr. Deborah C. Mash
Director

SHIP TO: Dr. Kinga Szigeti
University at Buffalo
Department of Neurology
875 Ellicott St., room 6069
Buffalo, NY 14203

DATE: August 21, 2013

Shipping Method	Shipping Terms	Payment Terms
N/A	Quick Stat Courier	30 days from receipt of specimens
Human Autopsy Confirmed- Alzheimer's disease and Normal Control (frozen specimens) *Annotated with clinical demographics*		

Regions of Interest	Quantity	Unit Price	Line Total
Alzheimer's disease - Temporal Cortex	250	$30.00	$7,500.00
Normal Control – Temporal Cortex	175	$30.00	$5,250.00
		TOTAL	**$12,750.00**

Please make check payable to: **University of Miami BRAIN BANK #330798**

NIAGADS The National Institute on Aging Genetics of Alzheimer's Disease Data Storage Site

DATE: October 4, 2013

TO: Kinga Szigeti, M.D., Ph.D.
University of Pennsylvania

FROM: DATA USE COMMITTEE

SUBJ: DUC APPROVAL FOR ACCESS - NG00022, 23, 24, 26, 29, 31

We are pleased to inform you that you have received NIA Genetics of Alzheimer's Disease Data Storage site (NIAGADS) Data Use Committee (DUC) approval to access the dataset with accession number NG00022, 23, 24, 26, 29 & 31. The requested data can only be used as prescribed in the approved application. Any major changes to your project must first be approved by the DUC.

This approval is good for one (1) year. Annual renewals must be submitted to NIAGADS by October 3rd, 2013 to retain access to the requested data. The data can be found here (note: you must be logged-in for the download links to be visible):

NG00022 - https://www.niagads.org/dataset/ng00022
NG00023 - https://www.niagads.org/dataset/ng00023
NG00024 - https://www.niagads.org/dataset/ng00024
NG00026 - https://www.niagads.org/dataset/ng00026
NG00029 - https://www.niagads.org/dataset/ng00028
NG00031 - https://www.niagads.org/dataset/ng00031

Please contact the NIAGADS office at (215) 898-9702 if you have questions or need assistance accessing the data.

Li-San Wang, Ph.D.
PI, NIAGADS
On behalf of the NIAGADS Data Use Committee

UHealth
UNIVERSITY OF MIAMI HEALTH SYSTEM
Psychiatry & Behavioral Sciences

MILLER
SCHOOL OF MEDICINE
UNIVERSITY OF MIAMI

Amanda J. Myers, PhD.
Associate Professor,
Department of Psychiatry & Behavioral Sciences,
Program in Neuroscience,
Interdepartmental Program in Human Genetics & Genomics
305-243-3522

http://labs.med.miami.edu/myers/
google scholar
researcherid
amyers@med.miami.edu

November 4, 2013

Dear Kinga,

It is my pleasure to support your grant entitled, “Narrowing the gap in the genetic architecture of Alzheimer’s disease” being submitted to the NIA as an R01application. As a consultant on the grant for years 4 and 5, I will assist with the replication study of specific aim 2. I have an extensive brain expression dataset with corresponding DNA samples. I will provide you with access to the gene expression data and assist in the locus specific genotyping of the corresponding DNA samples. I estimate to spend approximately 192 hours per year for years 4 and 5 of the grant. The hourly rate is 52 USD.

I am looking forward to working with you on this interesting project.

With warm regards,

AJ Myers

Amanda J. Myers
Associate Professor, Laboratory of Functional Neurogenomics
Department of Psychiatry, Miller School of Medicine

Perelman
School of Medicine
UNIVERSITY *of* PENNSYLVANIA

NIAGADS

Li-San Wang, Ph.D.
Assistant Professor
Department of Pathology and Laboratory Medicine
Penn Institute on Aging
Penn Institute of Biomedical Informatics (IBI)
Penn Genome Frontiers Institute (PGFI)
NIA Genetics of Alzheimer's Disease Data Storage Site (NIAGADS)
1424 Blockley Hall, 423 Guardian Drive
University of Pennsylvania Perelman School of Medicine
Philadelphia, PA 19104-6021
Phone: (215) 746-7015 Fax: (215) 573-3111
Email: lswang@mail.med.upenn.edu

September 30, 2013

Kinga Szigeti, MD. PhD.
D-2 Buffalo General Hospital
100 High Street
Buffalo, NY 14203

Dear Kinga,

I am writing this letter to express my emphatic willingness to provide assistance in the use of National Institute on Aging Genetics of Alzheimer's Disease Data Storage Site (NIAGADS) genetic and related neuropathology data to aid in your research in elucidating the contribution of copy number variation to the genetics of Alzheimer's disease. As we have discussed before, the proposed work is exciting and potentially very significant. Your proposal uses innovative strategies and will substantially contribute to the assessment of the role of CNVs in AD. Through the NIAGADS a large amount of data relevant to your proposal is available. I am glad to state that your application for the NIAGADS data has been approved, which provides access to the intensity data of the ADC 1-3 datasets and data access to the ROSMAP and MIRAGE studies are in progress.

I wish you much luck with this important application and look forward to continued, fruitful collaboration.

Sincerely,

LiSan Wang

Li-San Wang, Ph.D.
Assistant Professor of Pathology and Laboratory Medicine
University of Pennsylvania Perelman School of Medicine

University at Buffalo
The State University of New York

Gil I. Wolfe, M.D., FAAN
Irvin and Rosemary Smith Professor and Chairman
Department of Neurology
School of Medicine and Biomedical Sciences

October 21, 2013

Dr. Cathleen Cooper, Acting Director
Center for Scientific Review
Division of Receipt and Referral

RE: Kinga Szigeti, M.D., Ph.D. (Applicant)
Title of Grant: *Narrowing the gap in the genetic architecture of Alzheimer's disease*

Dear Dr. Cooper:

I am writing in strong support of the application of Kinga Szigeti, M.D., Ph.D. She has the full commitment from the Department of Neurology to carry out the research proposed in the above-mentioned grant.

Dr. Szigeti received her medical training at the University Medical School of Pecs, Hungary, and graduated with an M.D. summa cum laude degree in 1994. She pursued a postdoctoral research fellowship at that same institution from 1994-1997, and then continued her research at Harvard University from 1997-1998, focusing on the role of *ras* in PC12 cell differentiation. She then pursued and completed her neurology residency training at the State University of New York at Buffalo from 1998-2002, where she served as chief resident from 2001-2002. Dr. Szigeti subsequently studied as a fellow in clinical genetics (with an emphasis on neurogenetics) at Baylor College of Medicine from 2002-2004. Throughout this period she continued to work on her doctorate and in 2006, Dr. Szigeti received her Ph.D. from the University Medical School of Szeged, Hungary. In September 2006, she was appointed to the faculty at Baylor as Assistant Professor in the Departments of Neurology and Molecular and Human Genetics. She has been certified by the American Board of Psychiatry and Neurology in neurology.

Dr. Szigeti was recruited back to Buffalo in 2010 by the Dean of our Medical School, Dr. Michael Cain. She is recognized nationally and internationally for her expertise in the genetics of Alzheimer's disease, and is currently playing an integral role in the Alzheimer's disease research and clinical program in our department. Although Dr. Szigeti and I specialize in different areas of neurology, I know that senior investigators in Alzheimer's disease, such as Roger Rosenberg, my former colleague at Univ. of Texas Southwestern Medical Center, think very highly of her. She has successfully established an Alzheimer's Disease and Memory Disorder Center, which has filled an unmet need in Western New York.

Upon her move to Buffalo, Dr. Szigeti continued the established collaborations with the Department of Molecular and Human Genetics at Baylor College of Medicine. At the same time, she developed new collaborations in Buffalo integrating multiple specialties (neuropsychology, neuroimaging, ontology, genetics and clinical care) to further advance her dementia research.

As her Chairman, I am fully committed to allowing her to continue her focus on the development of her academic and research career. She is currently the only member of my faculty who has dedicated laboratory space in the University's new Clinical Translational Research Center (CTRC), which opened last year. The CTRC is a state-of-the-art research center located on the Buffalo-Niagara Medical Campus, with brand new laboratory space dedicated to only the best of the best. The fact that Dr. Szigeti was asked to join the CTRC is a testament to the support she has not only Department-wide, but also University-wide.

Dr. Szigeti currently holds a K23 Career Development Award and has completed a New Investigator Research Grant from the Alzheimer's Association. This RO1 application is a logical extension utilizing the novel analytical methods and technologies on a larger scale. Dr. Szigeti will continue to have access to her mentors, Norma Nowak, Ph.D., and James Lupski, M.D., Ph.D., at Baylor College of Medicine.

As her Chairman, I agree to release Dr. Szigeti from other clinical and administrative duties and activities so that she can devote at least 75% of her time for the performance of these grant activities. I am firmly committed to protect and maintain the time commitment necessary to conduct her research. Additionally, I will ensure that she continues to have adequate office space and support staff to fulfill the requirements set forth in the grant.

Thank you for considering Dr. Szigeti's grant proposal. She is a valued member of our faculty with outstanding credentials, and we are committed to her continued development as a productive, independent investigator. If I can be of further assistance, please don't hesitate to contact me.

Sincerely,

Gil I Wolfe / Gil

Gil I. Wolfe, M.D.
Irvin & Rosemary Smith Professor and Chair
Department of Neurology

ROSWELL PARK CANCER INSTITUTE

connecting for life

Elm & Carlton Streets
Buffalo, NY 14263
716-845-2300
www.roswellpark.org
E-mail: askrpci@roswellpark.org

October 31, 2013

Kinga Szigeti, MD, PhD
Assistant Professor of Neurology
Director, Alzheimer's Disease and Memory Disorders Center
University at Buffalo, SUNY

Dear Dr.Szigeti,

I am delighted to support your grant entitled, “Narrowing the gap in the genetic architecture of Alzheimer’s disease” being submitted for the NIH R01application. As your proposed collaborator, I feel I have the bioinformatics expertise to complement your strengths in bioinformatics and biostatistical data analysis needed in this study.

Although I have only been an Assistant Professor in the Department of Biostatistics and Bioinformatics at Roswell Park Cancer Institute (RPCI) a short time, I feel we have laid a strong foundation for this collaboration. I am currently the leader of the Bioinformatics Support for Center of Genomic Medicine at Roswell Park Cancer Institute. As you know, I have extensive experience working with genomic datasets including a number of large-scale microarray studies, RNA seq and many NGS studies. In addition, I also have experience developing bioinformatics software packages including CREST which identifies somatic structure variations from NGS data, PCAP, a parallel whole-gene assembly program, PolyFreq, a program to identify SNPs with allele frequency, ChiP-PAM, a program for ChiP-Seq data analysis using motif finding, and FUSIM, a simulation tool for fusion discovery from RNA-Seq data. My current research on high-density microarray and RNA seq data analysis and my past experience working on computational genomics, network biology and proteomics provide me with expertise necessary to assist you with your data analysis.

I am looking forward to continue working with you on this interesting project. Good luck with your proposal

Sincerely,

Jianmin Wang, PhD
Assistant Member, Biostatistics Department
Roswell Park Cancer Institute
Elm & Carlton Streets
Buffalo, NY 14263
Phone: (716) 845-1499
Email: Jianmin.Wang@roswellpark.org

ROSWELL PARK
CANCER INSTITUTE

Elm & Carlton Streets
Buffalo, NY 14263
716-845-2300
www.roswellpark.org
E-mail: askrpci@roswellpark.org

connecting for life

October 23, 2013

Kinga Szigeti, MD, PhD
Assistant Professor of Neurology
Director, Alzheimer's Disease and Memory Disorders Center University at Buffalo, SUNY

Dear Dr. Szigeti,

I am delighted to support your grant entitled, “Narrowing the gap in the genetic architecture of Alzheimer's disease” being submitted to the national Institute on Aging. As your proposed collaborator, I feel I have the biostatistics and bioinformatics expertise to complement your strengths in data analysis needed in this study.

As a senior biostatistician at Roswell Park Cancer Institute, I have the expertise and motivation necessary to successfully carry out the proposed work. I collaborated with Dr. Szigeti (PI of this proposal) and other investigators on several projects using cutting-edge statistics and genomics method to study the risk factors of Alzheimer disease and other genomic related diseases. I would provide statistical support on experimental design, clinical data and genomic data integration, and analysis the integrated data for the specified aims. I have the expertise and motivation necessary to successfully carry out the proposed work. With a broad background in bioinformatics, biostatistics, and population, clinical related statistical support, I have accumulated extensive experience in statistical and bioinformatics analysis, incorporating the conventional clinical data analysis and high-throughput genomics datasets including large-scale high-density microarray and second-generation sequencing studies.

I am looking forward to continue working with you on this interesting project. Good luck with your proposal.

Sincerely,

L Yan

Li Yan, PhD
Senior Statistician,
Biostatistics Department
Roswell Park Cancer Institute
Elm & Carlton Streets
Buffalo, NY 14263
Phone: (716) 845-7757
Email: Li.Yan@roswellpark.org

Matthew Huentelman, PhD
Associate Professor, Neurogenomics
Co-Director, Center for Rare Childhood Disorders
Translational Genomics Research Institute
445 N Fifth Street, 5th Floor
Phoenix, Arizona 85004

October 31, 2013

Kinga Szigeti, MD, PhD
Assistant Professor of Neurology
Director, Alzheimer's Disease and Memory Disorders Center University at Buffalo, SUNY

Dear Kinga,

I am writing this letter to confirm the support of the proposed CNV analysis. You already have access to the intensity data files of the TGEN Affymetrix 6.0 GWAS study and have been working with the data. The proposed work is exciting and has the potential to elucidate additional genetic players in Alzheimer's disease.

I wish you luck with your endeavors and stand at the ready to help you attain the goals outlined in your proposal.

Sincerely,

Matt Huentelman, PhD

Resource Sharing

Data sharing plan: I have familiarized myself with the sharing policy of the National Institute of Aging and I will adhere to the policy. We will release phenotype, copy number variation, and RNA sequence data from all parts of this study after removal of personal identifying information within 12 months of the completion of the study. We will develop a web-based tool to visualize the copy number calls to be able to share our data effectively. We will ask that other users of these data follow customary processes in which publication of secondary analyses is not undertaken until the main study results have been published by the data production team.

Sharing of model organisms: New model organisms will not be developed during this project.