The content of this application is copyrighted. Investigators and others may use the content only for nonprofit educational purposes provided the document remains unchanged and the Principal Investigator(s), the grantee organization(s), and the NHGRI are credited. All documents are in PDF format.

PI: **ROCHA ABECASIS, GONCALO**	Title: Analysis of Coding Variation and Macular Degeneration in 38,000 Individuals	
Received: 11/23/2011	FOA: PAR11-210	Council: 05/2012
Competition ID: ADOBE-FORMS-B1	FOA Title: CENTER FOR INHERITED DISEASE RESEARCH (CIDR) HIGH THROUGHPUT GENOTYPING AND SEQUENCING RESOURCE ACCESS (X01)	
1 X01 HG	Dual:	Accession Number: 3444460
IPF: 1506502	Organization: UNIVERSITY OF MICHIGAN AT ANN ARBOR	
Former Number:	Department:	
IRG/SRG: ZHG1 SRC (99)	AIDS: N	Expedited: N
Subtotal Direct Costs (excludes consortium F&A) Year 1: 0	Animals: N Humans: Y Clinical Trial: N Current HS Code: 20 HESC: N	New Investigator: N Early Stage Investigator: N
Senior/Key Personnel:	*Organization:*	*Role Category:*
Goncalo Abecasis	Regents of the University of Michigan	PD/PI
Rando Allikmets	Columbia University	Co-Investigator
Paul Baird	University of Melbourne	Co-Investigator
Chen-Yu Cheng	Singapore Eye Research Institute	Co-Investigator
Margaret DeAngelis	University of Utah	Co-Investigator
Jonathan Haines	Vanderbilt University	Co-Investigator
Micheal Klein	Oregon Health & Science University	Co-Investigator
Thierry Leveillard	Institut de la Vision Inserm	Co-Investigator
Sudha Lyengar	Case Western Reserve University	Co-Investigator
Margaret Pericak-Vance	University of Miami	Co-Investigator
Dwight Stambolian	University of Pennsylvania	Co-Investigator
Anand Swaroop	National Eye Institute	Co-Investigator
Jie Jin Wang	University of Sydney	Co-Investigator
Bernhard Weber	University of Regensburg	Co-Investigator
Dan Weeks	University of Pittsburgh	Co-Investigator
Tien Wong	National University of Signapore	Co-Investigator

Appendices

appendix_data_summary,appendix_roles,appendix_data_dictionar

SF 424 (R&R) APPLICATION FOR FEDERAL ASSISTANCE

15. ESTIMATED PROJECT FUNDING

a. Total Federal Funds Requested	0.00
b. Total Non-Federal Funds	0.00
c. Total Federal & Non-Federal Funds	0.00
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 penalities. (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

Add Attachment | Delete Attachment | View Attachment

19. Authorized Representative

Prefix: Mrs.   * First Name: Elaine   Middle Name: L.

* Last Name: Brock   Suffix:

* Position/Title: Sr.Associate Director,DRDA

* Organization: Regents of the University of Michigan

Department: Division:

* Street1: 3003 S. State St

Street2:

* City: Ann Arbor   County / Parish: Washtenaw

* State: MI: Michigan   Province:

* Country: USA: UNITED STATES   * ZIP / Postal Code: 481091274

* Phone Number: 734-764-7246   Fax Number:

* Email: tmax@umich.edu

* Signature of Authorized Representative	* Date Signed
Elaine.Brock	11/22/2011

20. Pre-application

Add Attachment | Delete Attachment | View Attachment

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

	Page Numbers
SF 424 R&R Face Page	1
Table of Contents	3
Research & Related Other Project Information	4
Project Summary/Abstract (Description)	5
Public Health Relevance Statement (Narrative attachment)	6
Research & Related Senior/Key Person	7
Biographical Sketches for each listed Senior/Key Person	16
PHS 398 Specific Cover Page Supplement	71
PHS 398 Specific Research Plan	73
Specific Aims	74
Research Strategy	75
Human Subjects Sections	81
Protection of Human Subjects	81
Women &Minorities	83
Planned Enrollment Table	84
Children	85
Bibliography & References Cited	86
Consortium/Contractual	89
Resource Sharing Plan	90

Appendix

Number of Attachments in Appendix: 3

RESEARCH & RELATED Other Project Information

1. * Are Human Subjects Involved? [X] Yes [ ] No
   - 1.a If YES to Human Subjects
   - Is the Project Exempt from Federal regulations? [ ] Yes [X] No
   - If yes, check appropriate exemption number. [ ] 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] 6
   - If no, is the IRB review Pending? [X] Yes [ ] No
   - IRB Approval Date:
   - Human Subject Assurance Number: 00004969
2. * Are Vertebrate Animals Used? [ ] Yes [X] 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 [X] No

4.a. * Does this project have an actual or potential impact on the environment? [ ] Yes [X] 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 [X] No

5.a. If yes, please explain:

6. * Does this project involve activities outside of the United States or partnerships with international collaborators? [ ] Yes [X] No

6.a. If yes, identify countries:

6.b. Optional Explanation:

7. * Project Summary/Abstract Abstract.pdf | Add Attachment | Delete Attachment | View Attachment

8. * Project Narrative Narrative.pdf | Add Attachment | Delete Attachment | View Attachment

9. Bibliography & References Cited References.pdf | Add Attachment | Delete Attachment | View Attachment

10. Facilities & Other Resources | Add Attachment | Delete Attachment | View Attachment

11. Equipment | Add Attachment | Delete Attachment | View Attachment

12. Other Attachments Add Attachments | Delete Attachments | View Attachments [ ]

ABSTRACT

Age-related macular degeneration (AMD) is a progressive neurodegenerative disease, which affects quality of life for millions of elderly individuals worldwide. In the United States, AMD accounts for the majority of cases of blindness. As the population ages and treatments for preventable blindness (such as cataract) become more widely available, AMD is expected to become the major cause of blindness worldwide.

Here, we propose to use new, cost-effective genotyping arrays developed with input from the AMD Gene Consortium members to systematically evaluate the contribution of rare protein coding variation to disease susceptibility. These arrays assess most of the variation that could be examined by sequencing but allow data to be generated more rapidly and for modest cost. Using a customized version of the Illumina exome chip to examine ~14,000 advanced AMD cases, ~6,000 cases of less advanced disease (all with large drusen) and ~18,000 geographically matched controls, we expect to identify genes where rare coding variants of clear functional impact are associated with disease and provide more direct insights into disease mechanisms.

NARRATIVE

In developed countries, age-related macular degeneration (AMD) is the most common cause of blindness in the elderly. The disease affects quality of life for millions of elderly individuals worldwide. Many regions of the DNA have been associated with disease, but the precise mechanisms that connect these genetic variants to disease remain unknown. Here, we propose to carry out more detailed genetic analyses of 38,000 individuals – including 12,000 cases of severe disease – to investigate the connections between genetic variation and disease in greater detail.

PHS 398 Cover Page Supplement

OMB Number: 0925-0001

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

Prefix:
* First Name: Goncalo
Middle Name:
* Last Name: Abecasis
Suffix:

2. Human Subjects

Clinical Trial? ☒ No ☐ Yes

* Agency-Defined Phase III Clinical Trial? ☐ No ☐ Yes

3. Applicant Organization Contact

Person to be contacted on matters involving this application

Prefix: Ms.
* First Name: Therese
Middle Name: G
* Last Name: Maxwell
Suffix:

* Phone Number: 734-764-7246
Fax Number:
Email: tmax@umich.edu

* Title: Project Representative

* Street1: Div Res Dev & Admin
Street2: 1040 Wolverine Tower
* City: Ann Arbor
County/Parish:
* State: MI: Michigan
Province:
* Country: USA: UNITED STATES
* Zip / Postal Code: 481091274

PHS 398 Cover Page Supplement

4. 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://stemcells.nih.gov/research/registry/. 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.

OMB Number: 0925-0001

PHS 398 Research Plan

1. Application Type:

From SF 424 (R&R) Cover Page. The response provided on that page, regarding the type of application being submitted, is repeated for your reference, as you attach the appropriate sections of the Research Plan.

*Type of Application:

☒ New ☐ Resubmission ☐ Renewal ☐ Continuation ☐ Revision

2. Research Plan Attachments:

Please attach applicable sections of the research plan, below.

Section	Attachment			
1. Introduction to Application (for RESUBMISSION or REVISION only)		Add Attachment	Delete Attachment	View Attachment
2. Specific Aims	SpecificAims.pdf	Add Attachment	Delete Attachment	View Attachment
3. *Research Strategy	ResearchPlan.pdf	Add Attachment	Delete Attachment	View Attachment
4. Inclusion Enrollment Report		Add Attachment	Delete Attachment	View Attachment
5. Progress Report Publication List		Add Attachment	Delete Attachment	View Attachment

Human Subjects Sections

Section	Attachment			
6. Protection of Human Subjects	HumanSubjects.pdf	Add Attachment	Delete Attachment	View Attachment
7. Inclusion of Women and Minorities	WomenAndMinorities.pdf	Add Attachment	Delete Attachment	View Attachment
8. Targeted/Planned Enrollment Table	EnrollmentTable.pdf	Add Attachment	Delete Attachment	View Attachment
9. Inclusion of Children	Children.pdf	Add Attachment	Delete Attachment	View Attachment

Other Research Plan Sections

Section	Attachment			
10. Vertebrate Animals		Add Attachment	Delete Attachment	View Attachment
11. Select Agent Research		Add Attachment	Delete Attachment	View Attachment
12. Multiple PD/PI Leadership Plan		Add Attachment	Delete Attachment	View Attachment
13. Consortium/Contractual Arrangements	Consortium.pdf	Add Attachment	Delete Attachment	View Attachment
14. Letters of Support		Add Attachment	Delete Attachment	View Attachment
15. Resource Sharing Plan(s)	ResourceSharing.pdf	Add Attachment	Delete Attachment	View Attachment

16. Appendix Add Attachments Remove Attachments View Attachments

SPECIFIC AIMS

Age-related macular degeneration (AMD) is a progressive neurodegenerative disease, which affects quality of life for millions of elderly individuals worldwide. In the United States, AMD accounts for the majority of cases of blindness. As the population ages and treatments for preventable blindness (such as cataract) become more widely available, AMD is expected to become the major cause of blindness worldwide.

Over the past several years, great strides have been made in our understanding of disease susceptibility. Common genetic variants in the complement pathway (near the *CFH*, *C3, C2*, *CFB* and *CFI* genes) and elsewhere (near *ARMS2/HTRA1*, *TIMP3*, *LIPC, CETP*) have been associated with disease risk. Most recently, our own ongoing studies in the AMD Gene Consortium have increased the number of known disease susceptibility loci to eighteen. Illustrating the principle that identification of additional genetic susceptibility loci gradually reveals the molecular pathways underlying disease susceptibility, we note that genomic studies of AMD at first highlighted the role of complement pathway in disease susceptibility, then of HDL-C associated variants and, most recently, of extracellular matrix proteins and regulators of angiogenesis (in AMD Gene Consortium unpublished data).

While current results already provide insights into the molecular mechanisms underlying AMD onset and progression, it is important to note that even when association has been robustly established, elucidation of precise molecular mechanisms remains challenging, requiring laborious and expensive follow-up using sequencing (to identify rare loss of function variants) and *in vitro* and model organism experiments (to clarify how altering function of the genes in each locus might impact susceptibility).

Here, we propose to use new, cost-effective genotyping technologies developed with input from the AMD Gene Consortium members to systematically evaluate the contribution of rare protein coding variation to disease susceptibility. These arrays, designed based on an analysis of sequence data for >12,000 individuals, assess most of the variation that could be examined by sequencing but allow data to be generated more rapidly and for modest cost. Using a customized version of the Illumina exome chip to examine ~14,000 advanced AMD cases, ~6,000 cases of less advanced disease (all with large drusen) and ~18,000 geographically matched controls, we expect to identify genes where rare coding variants of clear functional impact are associated with disease and provide more direct insights into disease mechanisms.

AIM 1: Customize the Illumina Exome Chip to Increase Coverage of Genes and Regions Implicated in AMD Susceptibility. Using publicly available data and sequence data generated by participating investigators, we plan to augment the content of the Illumina Exome Array to increase coverage of regions of the genome of the greatest interest for genetic studies of AMD. In addition, we plan to modify the array to ensure variants of highest interest are targeted with redundant primers whenever possible (minimizing the chances of assay failure).

AIM 2: Characterize patterns of rare genetic variation in 38,000 individuals, including 14,000 advanced AMD. Participating investigators have accumulated a large and informative set of AMD cases and controls. The combined set is well powered to enable genetic discoveries, providing strong support for ~19 genomewide significant loci (unpublished data). We propose to use our customized arrays to characterize coding variation across 38,000 individuals drawn from this set of phenotyped individuals. Before proceeding to detailed association analysis (aim 3), we will describe patterns of genetic variation across consortium samples and explore different strategies for genotype calling of rare variants and for controlling for stratification in case control studies of rare genetic variants.

AIM 3: Systematically evaluate the contribution of rare coding variants to disease susceptibility, within previously identified loci and across the genome. We will systematically assess the contribution of genetic variation to disease susceptibility in previously identified susceptibility loci, with the objective of cataloging likely functional variants at each locus, better understanding the biological mechanism through which genetic variation modifies disease susceptibility and improving our ability to predict disease outcomes. In addition to dissecting the biology of known loci, we will also scan the genome for new susceptibility loci, furthering our understanding of AMD genetics and biology.

To achieve these aims, our proposed project brings together a multi-disciplinary team of investigators with expertise in ophthalmology, cell and molecular biology, genetics and statistical genetics. We believe AMD is an ideal disease in which to deploy new genomic assays (such as the Illumina exome chip), and that – just like with early genomewide association studies of AMD – the results of these studies will inform the design of complex trait studies for years to come. To ensure that the data we generate is maximally useful, we will deposit generated data and analysis summaries into dbGap.

BACKGROUND AND SIGNIFICANCE

Age-related macular degeneration (AMD) is a progressive neurodegenerative disease, which affects quality of life for millions of elderly individuals worldwide. Clinical presentation of disease is diverse (KLEIN *et al.* 1997; KLEIN *et al.* 2002). In early stages, the disease is characterized by atrophy of the retinal pigment epithelium (RPE) and the appearance of large drusen, which are extracellular deposits of protein and lipid materials beneath the retina. In later stages, the disease is associated with extensive damage to the RPE and loss of photoreceptors. In these later stages, disease is usually classified as choroidal neovascularization (CNV) or geographic atrophy (GA). In CNV, or "wet" AMD, new blood vessels develop under the RPE. In GA, damage to the RPE and photoreceptors is accompanied by cellular and pigmentation alterations in the RPE and Bruch's membrane, but development of new blood vessels is absent (SWAROOP *et al.* 2009). Vision loss due to CNV can often be stabilized by intravitreal anti-angiogenic agents, whereas vision loss due to GA remains progressive and irreversible. In the United States, advanced AMD accounts for more than 54% of all blindness, impacting approximately 1.7 million people of age 40 years or older (FRIEDMAN *et al.* 2004). Another 7.3 million individuals have intermediate AMD (FRIEDMAN *et al.* 2004).

The contribution of genetics to AMD has been documented through reports of familial aggregation, of concordant phenotypes in twins, and of a higher risk of disease in first-degree relatives of affected individuals (HEIBA *et al.* 1994; KLAVER *et al.* 1998; SEDDON *et al.* 1997). Overall, it is estimated that first degree relatives of affected individuals exhibit 3 to 6 times higher risk of AMD. In the past several years, great progress has been made in identifying genetic loci that account for some of this familial aggregation. First, identification of a strong association between AMD and the *CFH* region prompted evaluation of genetic association between AMD and several other genes in the complement pathway (EDWARDS *et al.* 2005; HAINES *et al.* 2005; KLEIN *et al.* 2005). Several additional complement genes have now been strongly associated with AMD, including complement 2 (*C2*) and complement factor B (*CFB*) (GOLD *et al.* 2006), complement 3 (*C3*) (YATES *et al.* 2007), and complement factor I (*CFI*) (FAGERNESS *et al.* 2009). Outside the *CFH* region, the major genetic contributors to AMD risk identified to date reside in the *ARMS2*/*HTRA* region (DEWAN *et al.* 2006; JAKOBSDOTTIR *et al.* 2005; MALLER *et al.* 2006; RIVERA *et al.* 2005). More recently, genetic variants near *TIMP3* (a gene previously implicated in early onset maculopathy) and near *LIPC* and *CETP* have been associated with macular degeneration (CHEN *et al.* 2010; NEALE *et al.* 2010). Overall, currently identified variants are estimated to account for 40-60% of the genetic contribution to disease risk (GOLD *et al.* 2006; our own unpublished data).

The knowledge that has already been gained is substantial. It is now clear that the complement pathway plays an important role in disease development and progression; providing the impetus for many experiments aimed at establishing the precise mechanisms involved (RAYCHAUDHURI *et al.* 2011; WEISMANN *et al.* 2011). Variants in *LIPC* and *CETP* which are associated with both blood HDL-C levels and age related macular degeneration suggest interesting avenues for disease treatment and prevention; for example, some of the drugs initially developed to manage HDL-C levels and prevent heart disease may be effective in the treatment or prevention of AMD (CHEN *et al.* 2010). In addition to providing mechanistic understanding of disease processes, disease associated variants can be used together to predict which individuals are likely to develop AMD or progress from early disease manifestations to more severe forms of disease (CHEN *et al.* 2010; JAKOBSDOTTIR *et al.* 2009; SEDDON *et al.* 2009).

Still, many challenges remain. In many loci, the precise identity of the variants that cause disease is the subject of debate – for example, it is not clear whether HTRA1, ARMS2 or another protein mediates the effects of AMD associated genetic variants on chromosome 10 (DEWAN *et al.* 2006; FRITSCHE *et al.* 2008; KANDA *et al.* 2007) and there is an ongoing debate about whether copy number variants near CFH independently contribute to disease susceptibility (HUGHES *et al.* 2006; RAYCHAUDHURI *et al.* 2010). In addition, several of the loci identified so far appear to contain multiple independently associated susceptibility variants (GOLD *et al.* 2006; LI *et al.* 2006; MALLER *et al.* 2006; RAYCHAUDHURI *et al.* 2011; SIVAKUMARAN *et al.* 2011) so that until genetic variation is examined more systematically, it will be hard to fully quantify the total contribution of these loci to disease susceptibility.

The experiments proposed here will help address these questions. Specifically, we hope that by examining rare and common variation in known loci we will be able to more fully quantify the contributions of known susceptibility loci to disease and to identify the rare highly penetrant variants (particularly premature stop codons and splice site altering variants) that might clarify the molecular mechanisms linking known loci to disease. In addition, by systematically examining rare and common coding variants across the genome, we hope to identify additional susceptibility loci.

Several published results illustrate the possibilities and encourage us to be optimistic about the success of our planned experiments. For example, in an ongoing study of the genetics of blood lipid levels, we re-

sequenced the coding regions of genes identified in the first GWAS of blood lipid levels (KATHIRESAN *et al.* 2008; WILLER *et al.* 2008). Genotyping of variants discovered in these targeted experiments and of other nearby variants identified by the 1000 Genomes Project (THE 1000 GENOMES PROJECT 2010) doubled the variability in blood LDL cholesterol levels accounted for by these loci (SANNA *et al.* 2011). The example of IFIH1 and type 1 diabetes is also instructive. After association to the IFIH1 region was established using genomewide association (SMYTH *et al.* 2006), resequencing experiments and follow-up genotyping showed that several low frequency IFIH1 loss-of-function variants (a category which includes nonsense and splice altering variants) were twice as common in individuals without type 1 diabetes as in controls (NEJENTSEV *et al.* 2009) – suggesting that therapies blocking IFIH1 function might reduce the risk of type 1 diabetes. Obtaining similar information for *ARMS2*, *HTRA1*, and other genes in AMD susceptibility loci would help clarify the mechanisms underlying disease susceptibility in each locus. As a final example, we note that rare and highly penetrant variants have recently been used to clarify the role of *CFH* in AMD (RAYCHAUDHURI *et al.* 2011).

In summary, we expect our proposed studies to lead to improved understanding of existing disease loci, improved prediction of individual phenotypes, and the identification of new disease susceptibility loci. These advances will improve our knowledge of AMD biology and should ultimately facilitate the development of better strategies for disease treatment and prevention.

INNOVATION

The genomewide association studies of the past five years enabled the discovery of hundreds of new disease susceptibility loci for a variety of traits. Identification of these loci, together with follow-up experiments in mouse and model organisms, is already providing insights about disease biology (for examples, see MUSUNURU *et al.* 2010; SANKARAN *et al.* 2008).

Next generation sequencing studies, which access a broader class of genetic variants (including common and rare SNPs, but also short insertions and deletions and copy number variants) are expected to usher in a new wave of disease gene discovery, but remain laborious and costly. Our proposed research will pioneer the use of exome genotyping arrays, a new technology designed to bridge the gap between association studies of common variants and sequencing studies of rare variants. Consortium members have played key roles in the construction and design of the first generation exome arrays and in this proposal we propose to demonstrate how customization of the arrays can further enable the next generation of disease studies. Our customization will provide improved coverage of known disease susceptibility loci, enable fine-mapping of known loci and power discovery of disease loci by follow-up of GWAS findings.

Our results should have broad relevance not only in macular degeneration, but also for all diseases and across many disciplines: epidemiology, physiology, biology, medicine, and pharmaceutical science. We focus on AMD because genetic variants of large effect exist and contribute to disease susceptibility – maximizing our chances of success. The insights into disease etiology and new genetic risk factors to be discovered have the potential for transformative clinical impact. Further, we expect that the experimental approach and statistical and analytical methods we develop to carry out the experiments described here will be deployed in the study of a wide variety of traits.

APPROACH

A COLLABORATIVE TEAM OF COMPLEMENTARY STRENGTHS. We recognize that successful completion of the proposed studies will require a multi-disciplinary team. Towards this end, we have assembled a team with diverse strengths. Our team includes accomplished clinicians and ophthalmologists (Drs. Klein, Stambolian and Wong), experts in epidemiology (Drs. Cheng, Klein, Wang) and in gene mapping of ocular traits (Drs. Baird, Cheng, DeAngelis, Haines, Pericak-Vance), experts in statistical analysis of genetic association studies and in the development of related methods and software (Drs. Abecasis, Haines, Iyengar, Pericak-Vance and Weeks) as well as experts in the necessary follow-up functional experiments and interpretation (Drs. Allikmets, Leveillard, Swaroop, Weber).

THE AMD GENE CONSORTIUM. To accelerate discovery of susceptibility genes for age related macular degeneration, we formed the AMD Gene Consortium in 2010. In the past year, the consortium has successfully undertaken an imputation based meta-analysis including 7,650 cases of advanced AMD (defined as choroidal neovascularization (59%), geographic atrophy (26%) or both (15%)) and 51,812 controls (the number of controls is large mainly because Decode, with >30,000 controls participated in the analysis) and follow-up of 32 promising SNPs in >8,000 cases and >8,000 controls. Overall, the original meta-analysis included 14 different GWAS (each of which was first extended to include results for HapMap markers using genotype imputation (LI *et al.* 2010; MARCHINI *et al.* 2007)).

The meta-analysis examined 2,859,744 imputed and genotyped SNPs. Meta-analysis results were adjusted for population stratification using the genomic control method (DEVLIN and ROEDER 1999) (prior to adjustment λ_{GC}=1.05) and resulted in clear evidence for association in 19 regions of the genome (each with at least one SNP with $p < 5x10^{-8}$, see Table 1 below). Among these loci, 12 were previously associated with macular degeneration in GWAS studies (APOE, ARMS2/HTRA1, C2/CFB, C3, CETP, CFH, CFI, FRK/COL10A1, LIPC, SYN3/TIMP3, TNFRSF10A/LOC389641, VEGFA1) whereas 7 were newly identified in consortium analyses (COL8A1, IER3/DDR1, SLC16A8, COL15A1/TGFBR1, RAD51B, ADAMTS9, B3GALTL). We note that all loci, except APOE, retained $p < 5x10^{-8}$ in an age adjusted meta-analysis [which necessarily down-weights the contributions of samples where case and control ages are less well matched].

SNP	Chromosome, Position		Nearby Genes	EAF	Discovery		Follow-up		Joint	
					P	OR	*P*	OR	*P*	OR
Loci previously reported with genome-wide significance										
rs10490924/T	10	124.2 Mb	*ARMS2*	0.30	4×10^{-353}	2.7	2.8×10^{-190}	2.9	4×10^{-540}	2.8
rs10737680/A	1	195.0 Mb	*CFH*	0.64	1×10^{-283}	2.4	2.7×10^{-152}	2.5	1×10^{-434}	2.4
rs429608/G	6	32.0 Mb	*C2/CFB*	0.86	2×10^{-54}	1.6	2.4×10^{-37}	1.9	4×10^{-89}	1.7
rs2230199/C	19	6.7 Mb	*C3*	0.20	2×10^{-26}	1.4	3.4×10^{-17}	1.4	1×10^{-41}	1.4
rs5749482/G	22	31.4 Mb	*TIMP3*	0.74	6×10^{-13}	1.3	9.7×10^{-17}	1.4	2×10^{-26}	1.3
rs4420638/A	19	50.1 Mb	*APOE*	0.83	3×10^{-15}	1.3	4.2×10^{-7}	1.3	2×10^{-20}	1.3
rs1864163/G	16	55.6 Mb	*CETP*	0.76	8×10^{-13}	1.2	8.7×10^{-5}	1.2	7×10^{-16}	1.2
rs943080/T	6	43.9 Mb	*VEGFA*	0.51	4×10^{-12}	1.2	1.6×10^{-5}	1.1	9×10^{-16}	1.2
rs13278062/T	8	23.1 Mb	*TNFRSF10A*	0.48	7×10^{-10}	1.2	6.4×10^{-7}	1.1	3×10^{-15}	1.2
rs920915/C	15	56.5 Mb	*LIPC*	0.48	2×10^{-9}	1.1	0.0040	1.1	3×10^{-11}	1.1
rs4698775/G	4	110.8 Mb	*CFI*	0.31	2×10^{-10}	1.2	0.025	1.1	7×10^{-11}	1.1
rs3812111/T	6	116.6 Mb	*FRK/COL10A1*	0.64	7×10^{-8}	1.1	0.022	1.1	2×10^{-8}	1.1
Novel AMD loci										
rs13081855/T	3	101.0 Mb	*COL8A1*	0.10	4×10^{-11}	1.3	6.0×10^{-4}	1.2	4×10^{-13}	1.2
rs3130783/A	6	30.9 Mb	*IER3/DDR1*	0.79	1×10^{-6}	1.2	3.5×10^{-6}	1.2	2×10^{-11}	1.2
rs8135665/T	22	36.8 Mb	*SLC16A8*	0.21	8×10^{-8}	1.2	5.6×10^{-5}	1.1	2×10^{-11}	1.2
rs334353/T	9	100.9 Mb	*COL15A1/TGFBR1*	0.73	9×10^{-7}	1.1	6.7×10^{-6}	1.1	3×10^{-11}	1.1
rs8017304/A	14	67.9 Mb	*RAD51B*	0.61	9×10^{-7}	1.1	2.1×10^{-5}	1.1	9×10^{-11}	1.1
rs6795735/T	3	64.7 Mb	*ADAMTS9*	0.46	9×10^{-8}	1.1	0.0066	1.1	5×10^{-9}	1.1
rs9542236/C	13	30.7 Mb	*B3GALTL*	0.44	2×10^{-6}	1.1	0.0018	1.1	2×10^{-8}	1.1

Table 1 summarizes the results of the consortium GWAS meta-analysis and follow-up. Tabulated results include SNP id, risk allele, chromosome and position (according to NCBI genome build 36), nearby genes, risk allele frequency, and p-value and odds-ratios for meta-analysis of the GWAS samples, follow-up samples and all available data.

The results demonstrate our ability to successfully organize a consortium including a large fraction of the AMD cases and controls consented for genetic studies anywhere, to conduct successful meta-analyses and follow-up genotyping across the consortium, and the power of our combined consortium samples – which demonstrate clear association signals for previously identified disease susceptibility loci and enable identification of new loci.

ANALYSIS TEAMS AND GOVERNANCE STRUCTURE. The AMD Gene Consortium provides us with a governance structure and dedicated team of individuals, including representation of all the groups and institutions contributing samples to this experiment, to coordinate the interesting and challenging analyses that will result from our planned experiments. The consortium typically meets by teleconference every week, to discuss interim results, review proposed analysis plans, and distribute workload among interested investigators. A consortium steering committee, elected by consortium members, and including Drs. Gonçalo Abecasis, Margaret DeAngelis, Sudha Iyengar, Margaret Pericak-Vance, and Bernhard Weber meets once per month to help coordinate strategic direction for the consortium and (with input from other consortium members) help resolve scientific or logistic issues that arise. Analysis teams are formed around specific tasks, typically including volunteers from multiple interested groups, and develop analysis plans that are discussed and

reviewed on consortium weekly calls before implementation. For the GWAS meta-analysis and follow-up summarized above the analysis team included senior investigators Drs. Gonçalo Abecasis and Jonathan Haines as well as more junior investigators Drs. Iris Heid and Lars Fritsche (Regensburg University, Germany), Wei Chen (initially at the University of Michigan, now at the University of Pittsburgh), Matt Schu (Boston University), and Yi Yu (Tufts) as well as additional collaborators that contributed to specific analyses when needed.

We expect to follow this same governance model to organize the analyses of the data generated as part of this application. Specifically, the consortium will discuss analysis and publication priorities by teleconference, seek volunteers for project specific analysis teams that will develop these priorities into detailed analysis plans, discuss the analysis plan in consortium calls, and then implement these plans and work to quickly complete analyses, organize necessary follow-up experiments and publish results.

AIM 1: CUSTOMIZE EXOME CHIP FOR STUDIES OF AMD.

Using publicly available data and sequence data generated by participating investigators, we plan to augment the content of the Illumina Exome Array to increase coverage of regions of the genome of the greatest interest for genetic studies of AMD. In addition, we plan to modify the array to ensure variants of highest interest are targeted with redundant primers whenever possible (minimizing the chances of assay failure). Specifically, we plan to consult with CIDR to add up to 30,000 custom beads to the array, to:

(A) Ensure more complete coverage of coding variation in regions previously associated with AMD. Illumina genotyping arrays execute a single base extension assay for each assayed polymorphism. This assay involves placing a primer of ~60 bases adjacent to each SNP and extending this primer by a single base with a labeled nucleotide. For most SNPs, there is a choice of two 60 base primers (“forward” and “reverse”). The original array was designed to target variants observed at least 3 times (2 times for stop and splice variants) in ~12,000 sequenced whole genomes and exomes. For each targeted variant, a single primer was selected for the array based on considerations such as sequence uniqueness and the ability of the primer to successfully anneal to processed messenger RNA in addition to DNA (Dr. Abecasis, one of the contributors to the AMD Gene Consortium, was responsible for coding variant selection). We expect approximately 10-20% of selected assays will fail. To maximize array success, we propose to use two probes for every coding variant in the 19 regions associated with AMD, making it less likely that key assays will fail genotyping. In addition, we propose to augment SNP selection in these regions with lists of coding variants identified by consortium members. Consortium members have already sequenced >1,000 AMD cases and controls in a combination of targeted sequencing experiments targeting ~2.7Mb around GWAS loci, whole genome and whole exome sequencing experiments. By early 2012, we expect this number to rise to >3,000 samples. We expect this more detailed analysis (including additional variants and/or probes in the 1Mb around each of 19 signals or ~0.7% of the genome) will require ~3,000 additional probes. The basic array design includes ~300,000 probes.

(B) Enable Fine-Mapping of AMD Susceptibility Loci. Genotyping of all consortium samples with a custom array also offers opportunities to fine-map disease susceptibility loci (e.g. to identify the most strongly associated variants in each region, whether or not they are causal). To facilitate this, we propose to pick ~1,000 tag SNPs for each region using LD information and the linkage disequilibrium patterns observed in >1,000 individuals sequenced by the 1000 Genome Project to data. A similar tagging strategy, based on an earlier version of the 1000 Genomes Project data, improved coverage of each region from an average r^2 of ~0.45 – 0.55 (depending on GWAS array) to ~0.79 (based on metabochip genotyping) (VOIGHT *et al.* 2012). With 19 regions and our experience with design of the Illumina Metabochip arrays, we expect this analysis would require ~20,000-24,000 additional probes.

(C) Facilitate Assessment of Copy Number Variation. There is great interest in the possible contribution of copy number variants to AMD susceptibility (HUGHES *et al.* 2006; RAYCHAUDHURI *et al.* 2010; SIVAKUMARAN *et al.* 2011). Many of these coding variants can be genotyped by focusing on specific single nucleotide sequences that uniquely identify specific polymorphic alleles (SUDMANT *et al.* 2010) or by examining signal intensity at invariant probes (KORN *et al.* 2008; MCCARROLL *et al.* 2008). We propose to use ~3,000 probes to evaluate copy number variation in the 19 regions previously associated with AMD susceptibility. These will be selected to include both common single nucleotide probes that identify copy number (SUDMANT *et al.* 2010) and to evaluate signal intensity (KORN *et al.* 2008; MCCARROLL *et al.* 2008) using invariant probes in regions that overlap common copy number variants described by the 1000 Genomes Project.

(D) Further Follow-Up AMD GWAS Results. To facilitate identification of additional susceptibility loci, we plan to include tags for the top ~1,000 association signals in our consortium GWAS.

We expect that the additions suggested here (totaling approximately 30,000 extra beads and adding an estimated $5 - $10 in per array costs) represent a cost effective way of increasing the value of our proposed genotyping experiments. If our project is funded, we will work with NIH and the CIDR staff to identify an optimal set of additional content that maximizes scientific value and fits the available genotyping budget.

AIM 2: CHARACTERIZE RARE CODING VARIATION IN ~38,000 INDIVIDUALS.

Participating investigators have accumulated a large and informative set of AMD cases and controls. The combined set is well powered to enable genetic discoveries, providing strong support for 19 genomewide significant loci. We propose to use customized exome arrays to examine coding variation across ~14,000 advanced AMD cases, ~6,000 AMD cases with large drusen, and ~18,000 matched controls drawn from this set of phenotyped individuals. Before proceeding to detailed association analysis (aim 3), we will explore different strategies for genotype calling of rare variants and attempt to obtain the highly accurate genotypes.

ADDITIONAL QUALITY CONTROL: To evaluate performance, it is critical to have samples characterized with high accuracy using alternative platforms. Thus, we plan to work with CIDR to identify an appropriate set of samples (such as 1000 Genome samples) for which sequence data or family information is available and which will be included in our genotyping experiment. We expect these to account for ~1% of all samples (approximately 1 sample per 96 well plate). Good choices might include the deeply sequenced 1000 Genomes Project trios, which have been sequenced to >100x depth using a combination of 5 sequencing platforms (Illumina Genome Analyzers, Illumina HiSeq, SOLiD, Complete Genomics, and 454) and thus provide a useful yard stick for evaluating the accuracy of genotype calls generated using our own data, alternatives include using samples for which consortium members have generated sequence data.

GENOTYPE CALLING: Illumina exome arrays target variants that are much rarer than those in previous genotyping arrays – which will pose unique challenges to genotyping calling algorithms. In particular, it may be that algorithms that try to define genotypes by examining signal intensities for all individuals, one probe at a time (such as the default GenomeStudio algorithms or Illuminus (TEO *et al.* 2007)) will find these rare variants particularly challenging to call. Alternative algorithms, that examine signal intensities for all probes in each individual (such as GenoSNP (GIANNOULATOU *et al.* 2008) or OptiCall (http://bitbucket.org/tss101/opticall/)). To evaluate trade-offs between different genotyping algorithms, we plan to allocate ~1% of our genotyping capacity to trios and previously sequenced samples. In this way, we should be able to evaluate different calling algorithms and evaluate which algorithms produce the best concordance with sequence based genotypes and result in transmission rates of ~50% for rare variants. For variants assayed with multiple probes (see Aim 1A, previous page), we will investigate whether one probe or both should be used to obtain the most accurate possible genotypes.

SUMMARY: We expect that transforming genotyping assay results into accurate genotypes will be challenging, particularly for very rare variants. Thus, we plan to initially apply a variety of state-of-the-art genotyping algorithms to our data (GenomeStudio, Illuminus, GenoSNP and optiCall) and evaluate the quality of the resulting genotypes – by comparison with sequence data for the same samples and by examining evidence for transmission distortion among rare variants (which is an indication of genotyping error (MITCHELL *et al.* 2003)). We will also describe overall patterns of variation (including allele frequency spectrum and rates of variation) in genes in AMD associated regions and compare these to the rest of the genome.

AIM 3: RARE VARIANT ASSOCIATION ANALYSES

SUMMARY: The first focus of our association analyses will be to improve our understanding of existing AMD susceptibility loci (for example, by identifying associated rare variants that clarify which genes in a locus, when disrupted, can change risk of macular degeneration). Toward this end, we plan to conduct targeted analyses that include only genes in previously implicated regions and therefore require a more lenient Bonferroni significance threshold. A second goal will be to identify new AMD susceptibility loci that may have been missed in analyses focused on more common variants and to identify loci that are associated with disease progression or between neovascular and geographic atrophy.

INITIAL SINGLE VARIANT ANALYSES: We will first carry out standard single-SNP tests of association based on logistic regression under an additive genetic model. Additional tests, that take into account overall variant-load within coding regions of genes are described in the next section. We will monitor possible population stratification or other model misspecification and correct for it if necessary using standard methods including principal components of ancestry (PRICE *et al.* 2006), genomic control (DEVLIN and ROEDER 1999) and variance component models of relatedness between samples (KANG *et al.* 2010). We will evaluate whether corrections for population structure should be carried out separately for common and rare variants.

Our preference will be to analyze each of the individual sample collections separately and meta-analyze results, so as to reduce the possibility of inadvertently introducing confounding.

Given 14,000 advanced AMD cases and 18,000 controls and a p-value of threshold of $5x10^{-7}$ (corresponding to ~100,000 independent tests across the exome, which is the number of coding variants with frequency > 0.001 included in the exome chip), we expect 80% power to detect variants associated with an increased risk of ~2.8 per allele and a frequency of 0.001 or greater. Similar power will require larger effect sizes and/or allele frequencies in analysis of disease subtypes. Common AMD associated alleles in CFH and near HTRA1/ARMS2 and C2/CFB already exhibit large relative risks per allele (see Table 1). Given additional observations that rare functional variants can exhibit large effects (COHEN *et al.* 2004; RAYCHAUDHURI *et al.* 2011; SANNA *et al.* 2011) we believe expectations of a well-powered study are realistic. In fact, we expect AMD will make an ideal test case of the utility of exome focused genotyping assays.

A subset of our samples includes related individuals. To incorporate genotypes from related individuals, we will fit parametric models of association using the LAMP program developed by co-investigator Mingyao Li (LI *et al.* 2005). The LAMP program estimates a disease allele frequency, a set of allele frequencies and three penetrances (constrained by disease prevalence) for each SNP using all available genotype data, phenotype data and family structures and can accommodate both families and unrelated individuals in a joint analysis. To identify independently associated sets of variants in each locus, we will use stepwise logistic regression.

AGGREGATE ANALYSIS OF LOW FREQUENCY VARIANTS: Most genetic variants on the exome arrays are rare so that it will not be practical to individually assess their contribution to macular degeneration. Instead, we will search for groups of disease associated low-frequency variants that can be identified by searching for unusual clustering of genetic variation in these genes, particularly in comparisons of aggregate sets of variants between cases and controls. Most promising for low-frequency variants are "burden" tests, designed to detect association signal distributed over multiple variants. Michigan postdoctoral fellow Bingshan Li developed one of the first such tests (LI and LEAL 2008); several other tests are now available (e.g. MADSEN and BROWNING 2009; PRICE *et al.* 2010; WU *et al.* 2010; ZAWISTOWSKI *et al.* 2010). These and other related tests aggregate variants across a region with or without differential weighting of variants based on minor allele frequency and/or genome annotation. Our current favorites are the Wu method, which allows low-frequency variants at a locus to increase and/or decrease risk, and the Zawistowski method, which allows explicitly for genotype uncertainty. This is an active area of research and we will monitor developments.

Critical issues for these burden tests include how best to choose the regions across which to aggregate and which variants to include. Our initial analyses will focus on variants clustered by gene or by gene function (as annotated in the Gene Ontology database). We will assign different weights to different classes of variants by considering their likelihood of affecting protein structure or function. We will include similar rankings for insertion deletion polymorphisms and copy number variants. We will test whether the non-synonymous to synonymous SNP ratio for a given genomic segment differs between cases and controls. For the most interesting genes, we will contact collaborators to organize follow-up sequencing experiments.

Given an effective sample size of ~14,000 cases and 18,000 controls, an a p-value threshold of $2.5x10^{-6}$, corresponding to Bonferroni adjustment for testing of 20,000 independent groups (e.g. one per gene), we expect 80% power to detect association in genes where the cumulative frequency of rare variants is 1% and their average contribution to disease risk corresponds to a relative risk of >1.5. Again, larger contributions to disease risk are known for common variants at several AMD susceptibility loci and our design will also allow well powered comparisons between advanced AMD cases and individuals with less severe forms of disease.

OVERALL SUMMARY: Here, we request genotyping with a customized Illumina exome array of ~14,000 cases of advanced AMD (~60% with neovascular disease, ~25% with geographic atrophy and 15% with both), ~6,000 cases of less advanced disease (all with large drusen) and of ~18,000 matched controls (see Table 2 below). The samples are drawn from 24 different collections (each of which includes both cases and controls collected in a geographically restricted area at a ratio of between 2 cases per control and 2 controls per case). Following the advice of Dr. Hemin Chin (National Eye Institute, NIH), we plan to deposit genotypes and summary phenotypes into dbGap so as to maximize the value of these data. We currently expect that this will be possible for nearly all samples. Because of consent restrictions, for ~5,000 samples, we might be permitted to deposit only summary level data.

Sample Set	Number of Samples	DNA Source	Service Requested
AMD Cases, Advanced	14,000	Blood (All genotyped previously)	Exome Array + 30,000 custom beads
AMD Cases, Large Drusen	6,000		
Controls	18,000		

Table 2 summarizes the genotyping requested in this application.

PROTECTION OF HUMAN SUBJECTS

Risks to the Subjects

Human subjects involvement and characteristics

The involvement of human subjects in this research will be limited to the use of existing DNA samples and phenotype data. These samples will include individuals of both sexes, with a small excess of females – consistent with other samples of older individuals. Our genetic studies of age-related macular degeneration have previously received human subjects approval and review from each participating institute, including the University of Michigan. All study participants provided written informed consent.

To minimize the risk of revealing participant identities, we will deposit into dbGaP information only on a select set of phenotypes. At a minimum we expect these to include disease grade, age at ascertainment, sex, and smoking status. We will discuss with program staff whether to deposit additional phenotypes, although integrating such data across all participating studies may prove challenging.

Sharing of de-identified materials is covered by existing informed consents.

Sources of materials

Research material is available in the form of existing DNA samples and phenotype data from our participant studies. DNA will be used for genotyping. Phenotype information will be used for a wide range of statistical analyses. This information has been gathered by participating investigators as part of ongoing research, and so no new clinical information or samples are being collected for this project. All study samples are anonymized. Personal identifiers will not be available for any of the analyses described here.

Potential risks

DNA analysis could reveal information about genetic background that is not already known to the subjects. Because data are archival, the only additional risk to participants is the possibility of breach of confidentiality, which we will guard against carefully using secure data systems.

Adequacy of protection against risks

Recruitment and informed consent

All study data are archival. Data only for those subjects who provided written informed consent will be used for data analysis.

Protection against risk

Study personnel take every reasonable measure to ensure confidentiality for all study participants and data. The link between study ID and personal information will only reside with the original study investigators; each study will protect their information using password protected secure systems. Transmission of data between the different study sites will be accomplished using encrypted file transfer protocols. Database and individual file permissions are restricted will be restricted to the smallest set of people feasible.

DNA analysis results are for research purposes only and will not be shared with study participants. As noted above, all data will be stored in secure database systems. No personally identifiable information will be released to third parties.

Potential benefits of the proposed research to the subjects and others

There is no expected direct benefit to participants. The potential benefit to be derived by humankind is the identification of genetic variants that may predict disease outcomes and facilitate the development of better treatments. Eventually, these findings may allow clinicians to identify individuals at risk for the disease and lead to early or improved interventions that could reduce individual disease risk. Information on genetic variation also may lead to new drugs for improved treatment of macular degeneration. The potential benefits to be gained from these studies warrant the minimal risks to study participants.

In the unlikely event that genotype results lead to information that might be clinically actionable by the participants (e.g. diagnosis of high-risk genotype for dyslipidemia that might benefit from treatment with statins), we will consult with consortium ethicists and clinical experts to decide on course of action which might include notifying the relevant physician that additional testing may be relevant or noting that notification of attending physician and/or participant is forbidden by the consent.

Importance of the knowledge to be gained

The identification of genetic variants associated with macular degeneration has implications for both patient care and for better understanding basic biology. Identifying susceptibility variants may provide a means for clinicians to identify those patients at highest risk of disease. This would then allow for early intervention which could significantly reduce the risk of disease. Identification of susceptibility variants could also allow pharmaceutical companies to develop new compounds that target these variants.

Data and safety monitoring plan

This is not a randomized clinical trial. No direct subject accrual or testing is occurring. As noted above, all samples and data are anonymized. Thus, no external Data Safety Monitoring Board is planned.

INCLUSION OF WOMEN AND MINORITIES

Women comprise ~60% of all current study samples. No specific efforts were made preferentially to include or exclude women or men. In each collection site, the sample make-up largely reflects the local patient population. African-American individuals are under-represented in the overall study sample because disease is relatively rare among them.

TARGETED / PLANNED ENROLLMENT

Targeted/Planned Enrollment Table

This report format should NOT be used for data collection from study participants.

Study Title: Sequencing Study of Age Related Macular Degeneration

Total Planned Enrollment: 38,000

TARGETED/PLANNED ENROLLMENT: Number of Subjects			
Ethnic Category	**Sex/Gender**		
	Females	**Males**	**Total**
Hispanic or Latino	900	700	1,600
Not Hispanic or Latino	20,100	16,300	36,400
Ethnic Category: Total of All Subjects *	21,000	17,000	38,000
Racial Categories			
American Indian/Alaska Native	0	0	0
Asian	1,800	1,400	3,200
Native Hawaiian or Other Pacific Islander	0	0	0
Black or African American	275	225	500
White	18,925	15,375	34,300
Racial Categories: Total of All Subjects *	21,000	17,000	38,000

* The "Ethnic Category: Total of All Subjects" must be equal to the "Racial Categories: Total of All Subjects."

INCLUSION OF CHILDREN

No children will be included in this study.

CONSORTIUM/CONTRACTUAL ARRANGEMENTS

The AMD Gene Consortium executive committee will coordinate collaborative analyses of the data resulting from this application. The committee includes Drs. Abecasis, DeAngelis, Iyengar, Pericak-Vance, Weber as voting members. The executive committee will aim to achieve a consensus plan, collating input from all consortium members. In the unlikely event this is not possible, decisions will be made by majority vote.

An appendix to the application lists AMD Gene Consortium members in the executive committee as well as other members who expressed interest in participating in this specific experiment in a leading role.

RESOURCE SHARING PLAN

We pledge to meet or exceed all relevant NIH standards for data sharing. Specifically, we plan to deposit genotype data generated by CIDR, together with the phenotypes proposed for analysis in this application and appropriate covariates, into dbGaP. Prior to submission, we ascertained from all participating investigators whether genotype data for their samples could be deposited in dbGap and expect to be able to deposit individual genotypes and phenotypes for ~33,000 samples into dbGap. For the remaining ~5,000 samples we are consulting with the local IRBs about whether deposit into dbGap is possible. If deposit into dbGap is not possible, we will consult with program staff as to whether to exclude those samples from this proposed experiment or to deposit summary level data.