Persons using assistive technology may not be able to fully access information in this file. For assistance, e-mail niddk-cr@imsweb.com. Include the Web site and filename in your message.

FIND Data Archive

The FIND Archive contains the study protocol, manual, forms; data collected by the study; and analysis files. The files are organized into the following directories:

- Documentation
- Forms
- Data
- Dataset_Integrity_Check

1. Documentation

The Documentation directory contains the MOOP and the Protocol subdirectories for the main study and the eye sub-study.

1.1 Manual of Operation (MOOP)

The MOOP folder contains the MOOP for the main study and the eye sub-study:

- The "Main_Study" subdirectory contains the MOOP for the main study [The MOO 2.18.2010 no Certificate.doc]
- The "Eye_Study" subdirectory contains the MOOP for the eye sub-study [Eye MOO 3.01 21Jun2004.pdf] and its ten appendices in separate files.
    - FIND_Eye_MOO_Appendix_list.doc
    - App 1 Modified7StdF 1.01 20Feb2002.doc
    - App 2 AttachmentB 2.04 21Jun2004.pdf
    - App 3 Photography procedure for non PICs 1.04 21Jun2004.pdf
    - App 4 T. Log Eye Exam Invoice Form 2.02 14Jun2002.doc
    - App 5 Collaborating MD Appt Packet 1.02 14Jun2002.doc
    - App 6 Collaborating MD Synopsis 1.04 21Jun2004.pdf
    - App 7 T. Log, Eye Exam, Invoice-Chart 1.02 21Jun2004.pdf
    - App 8 Collaborating ECP-Chart Packet 1.02 21Jun2004.pdf
    - App 9 Collaborating ECP Synopsis-Chart 1.02 21Jun2004.pdf
    - App 10 NIDDK Phoenix Data 1.02 08Jul2004.pdf

1.2 Protocol

The Protocol folder contains the publication, Knowler_2005.pdf, which describes the experiment design and method and which is used as a proxy for the protocol for the main study and the sub-studies.

2. Forms

The Forms directory contains two forms used in the main study:

- The FIND Medical Questionnaire [MEDICAL_QUESTIONNAIR.doc]
- The FIND Medical Record Review [MEDICAL_RECORD_REVIEW.doc]

3. Data

The FIND Study has two types of subject collections: sib pair collections and case-control collections. The main study collected data about kidney phenotypes. An additional sub-study (the eye sub-study) collected data about retinopathy phenotypes from a subset of individuals. The FIND data is organized in 2 subfolders. For the main study, the raw data collected by the study forms is listed under "Form_based_data" subfolder, and the genetic analyses carried out using samples from FIND are listed under the "Analysis_Data" subfolder, which includes: linkage studies in two phases (CIDR I and CIDR II) using a partially overlapping set of samples, two MALD (Mapping by Admixture Disequilibrium) studies on different sample sets, a Genome-Wide Association Study (GWAS), and a small genotyping dataset using Perlegen platform.

For the linkage studies, the genome scan was performed in two phases, and the datasets are grouped by phases in folders named "CIDR_I" and "CIDR_II" respectively. Study phase II has approximately three times as many participants as phase I, and was genotyped on a different genome-wide marker scan (approximately 6,000 single-nucleotide polymorphisms (SNPs) vs. approximately 400 microsatellite markers).

Note: the American Indian (AI) population and some study subjects from other populations did not give permission to share their data with outside investigators; therefore the data from these subjects is not included in this data archive.

3.1. Main_Study

3.1.1. Form_based data

The raw data collected by the study forms contains the following data files:

- FIND_MQ_21Sep2011.csv – raw data collected by the Medical Questionnaire form
- FIND_MRR_21Sep2011.csv - raw data collected by the Medical Record Review form
- FIND_LabResults_21Sep2011.csv – raw lab test results
- FIND_PED_21Sep2011.csv – pedigree file for the FIND subjects
- FIND_Retin_Diagnosis_Results_21Sep2011.csv – diagnosis results in the Retinopathy study
- FIND_Retin_Eye_Exam_Analysis_21Sep2011.csv – Eye exam analysis in the Retinopathy study
- FIND_Retin_Eye_Exam_Results_21Sep2011.csv – Eye exam results in the Retinopathy study
- Data_Dictionary_21Sep2011.xls – data dictionary

3.1.2. Analysis data

3.1.2.1 CIDR_I

The analysis datasets in this folder contain the results from genomewide model-free linkage analysis,

using Haseman-Elston (H-E) regression, for diabetic nephropathy (DN) and related phenotypes as presented in Figures 1 and 3 of Iyengar et al [1] and in Figure 1 of Schelling et al [2]. It contains three supporting files and subfolders for the datasets:

- **README.doc**: A description of the data files.
- **FIND_CIDR_I_phenotypes.txt**: phenotype data
- **Mfd8_ersatz.txt**: A tab-delimited text file with genetic map positions of the Research Genetics screening set of microsatellite markers, which is based on the Marshfield screening set 8
- **ACR_SIBPAL**: Linkage analysis results for albumin/creatinine ratio (ACR)
- **DN_SIBPAL**: Linkage analysis results for diabetic nephropathy (DN)
- **GFR_SIBPAL**: Linkage analysis results for glomerular filtration rate (GFR).

3.1.2.2 CIDR_II

Phenotype_data

- **README_CIDR_II.doc**: description of these data files
- **FIND_CIDR2_Codebook.xls:** Guide to the data fields in the CIDR II files
- **FIND_CIDR2_ForRepository_30Sep2010.xls:** phenotype data
- **CIDR_II_Pedigrees.xls:** Pedigree structure information, proband status and sharing status
- **CIDR_II_DM_Definitions.doc:** Compound Phenotype Definitions for Diabetes Mellitus (DM) Criteria for FIND CIDR II

Genotype data

- **README_CIDR_II_Genotypes.doc**: description of these data files.
- **Illumina_IV.map**: Map file containing information on SNPs in the Illumina IV panel, on which the CIDR II data were typed
- **Chr*.ped:** Genotype data listed by individual chromosomes

3.1.2.3 MALD

The analysis datasets in this folder contain analysis datasets of the MALD (Mapping by Admixture Disequilibrium) studies.

MALD_AA subdirectory contains analysis datasets of the MALD studies on African American (AA) population.

- MALD_AA.geno.121710.csv – Genotype data file
- MALD_AA.pheno.041511.csv – Phenotype data file
- Readme.docx – description file for the genotype and phenotype data

MALD_MA subdirectory contains analysis datasets of the MALD (Mapping by Admixture Disequilibrium) studies on Mexican American (MA) population.

- MALD_MA_genotype_04012011.csv – Genotype data file
- MALD_MA_genotype_map_04012011.csv – map file for the genotype data
- MALD_MA_phenotype_04012011.csv – Phenotype data file
- Readme.doc – description file for the genotype and phenotype data

3.1.2.4 Genome-Wide Association Study (GWAS) data

The GWAS datasets include the phenotypic data for the GWAS performed using Affymatrix chip (932,535 SNP) on 2,621 subjects, and the genotype data will be released in dbGaP.

- FIND_Subject_PhenotypeDS.txt – phenotype data
- FIND_Subject_PhenotypeDD.xls – data dictionary for the phenotype data

3.1.2.5 Perlegen-Phase2 data

A pooled GWAS was performed using 4 case and 4 control pools of Mexican American descent. The data associated with creation of the case and control pools, as well as analysis, cannot be shared because it contained data from individuals who did not consent to share their data and it was not possible to dissociate the data for sharing. Loci with the best p-values, as well as loci used to adjust for population structure, were advanced to individual genotyping (N=923 samples).
This data (designated Phase 2) from the Perlegen pooled GWAS study is included in this release. The data release is limited to records of individuals who consented to general data sharing only.

- Phenotype.txt – phenotype data
- genotype.txt – genotype data
- snp_info.txt – SNP map data
- Sample QC_procedures.doc – QC procedures
- sample_QC_report.doc – QC report

3.2 **Eye substudy**

The Eye substudy analysis dataset includes the phenotypic of the retinopathy phenotypes for the additional sub-study, the "Eye Study" described in the Protocol. This study is published in the Investigative Ophthalmology & Visual Science 2008 by Arar et al. It includes the following:

- FIND Retinopathy README.doc – dataset description file
- FIND_Retinopathy_Key.xls - describes each data field, and lists the possible values of categorical variables
- FIND_combined_retinopathy.txt – contains study data
- FIND_combined_retinopathy_meds.txt – contains medications data

3.3 GAD

The gad.sas7bdat dataset contains the GAD antibody titers dataset.

4. Dataset_Integrity_Check

The FIND Dataset Integrity Check directory contains a report of an examination of the repository by statisticians and quality control specialists at the Repository. Datasets were checked for completeness, consistency, and usability. The published data from the FIND study were compared to values recalculated from the FIND data in the NIDDK repository.

The integrity check (DSIC_FIND_CIDRI_Diabetes_2007.pdf) is an analysis by repository statisticians using the dataset under "/Data/Main_Study/Analysis_Data/CIDR_I/ FIND_CIDR_I_phenotypes.txt" in this distribution against the study data published in Iyengar_Diabetes_2007.pdf [1].

The integrity check (DSIC_FIND_CIDRII_AJN_2011.pdf) is an analysis by repository statisticians using the "FIND_CIDR2_ForRepository_30Sep2010.xls" and "CIDR_II_Pedigrees .xls" data files under "/Data/ Main_Study/Analysis_Data/CIDR_II/Phenotype_data/" in this distribution against the study data published in Igo, et al [3].

References:

[1] Iyengar SK, Abboud HE, Goddard KA, Saad MF, Adler SG, Arar NH, Bowden DW, Duggirala R, Elston RC, Hanson RL, Ipp E, Kao WH, Kimmel PL, Klag MJ, Knowler WC, Meoni LA, Nelson RG, Nicholas SB, Pahl MV, Parekh RS, Quade SR, Rich SS, Rotter JI, Scavini M, Schelling JR, Sedor JR, Sehgal AR, Shah VO, Smith MW, Taylor KD, Winkler CA, Zager PG, Freedman BI; Family Investigation of Nephropathy and Diabetes Research Group. Genome-wide scans for diabetic nephropathy and albuminuria in multiethnic populations: the family investigation of nephropathy and diabetes (FIND). Diabetes. 2007 Jun;56(6):1577-85.

[2] Schelling JR, Abboud HE, Nicholas SB, Pahl MV, Sedor JR, Adler SG, Arar NH, Bowden DW, Elston RC, Freedman BI, Goddard KA, Guo X, Hanson RL, Ipp E, Iyengar SK, Jun G, Kao WH, Kasinath BS, Kimmel PL, Klag MJ, Knowler WC, Nelson RG, Parekh RS, Quade SR, Rich SS, Saad MF, Scavini M, Smith MW, Taylor K, Winkler CA, Zager PG, Shah VO; Family Investigation of Nephropathy and Diabetes Research Group. Genome-wide scan for estimated glomerular filtration rate in multi-ethnic diabetic populations: the Family Investigation of Nephropathy and Diabetes (FIND). Diabetes. 2008 Jan;57(1):235-43.

[3] Igo RP Jr, Iyengar SK, Nicholas SB, Goddard KA, Langefeld CD, Hanson RL, Duggirala R, Divers J, Abboud H, Adler SG, Arar NH, Horvath A, Elston RC, Bowden DW, Guo X, Ipp E, Kao WH, Kimmel PL, Knowler WC, Meoni LA, Molineros J, Nelson RG, Pahl MV, Parekh RS, Rasooly RS, Schelling JR, Shah VO, Smith MW, Winkler CA, Zager PG, Sedor JR, Freedman BI. Genomewide Linkage Scan for Diabetic Renal Failure and Albuminuria: The FIND Study. Am J Nephrol. 2011 Mar 31;33(5):381-389.