

## **GOKIND Data Archive Description**

The GOKIND data archive contains the study protocol, manual, procedures, study forms; data collected by the study; analysis files, and other descriptive documentation.

The archive is organized into the following directories:

- GOKIND\_Documents
- GOKIND\_Forms
- GOKIND\_Data

### **1. GOKIND Documents**

The “GOKIND\_Documents” directory contains the following documents:

- The “GOKIND MOOP” directory contains 19 files, which comprise the Manual of Operations in 17 pdf and 2 gif files. The “JDRF09\_genotyping” subdirectory contains the CDC instructions for genotype determination.
- The “GOKIND Protocol” directory contains the GOKIND protocol (36 pages) as a pdf file.
- GoKinD\_Integrity\_V2.pdf – A report describing the results of an analysis by repository statisticians and quality control specialists to ensure that the dataset distributed by the repository is a true copy of the study data.

### **2. GOKIND Forms**

The “GOKIND\_Forms” subdirectory contains version 2 of the 22 Phase 1 study forms in pdf format. See Appendix 1 for a list of the GOKIND forms.

### **3. GOKIND Data**

The GOKIND Data directory contains data and analysis files. The files are organized into the following subdirectories:

- GOKIND\_GAIN\_2006
- GOKIND\_Reports\_Trios\_Probands
- GOKIND\_SAS\_Data
- GOKIND\_SAS\_to\_Excel
- GOKIND\_SAS\_Variables
- GOKIND\_Supplemental\_Data

#### **3.1. GOKIND GAIN 2006**

The directory “GOKIND GAIN 2006” contains information describing the GOKIND participants that were excluded from the Genetic Association Information Network (GAIN) after review by the Technical Advisory Group (TAG). Of the 3119 total participants, 1795 went to GAIN and 1324 were excluded from GAIN.

The directory contains three files:

- Cover-Lette-ext2r.doc – A description of the data files
- rti\_famn.sas7bdat - SAS data file, 23 variables, 3119 records
- gk\_ext1.rti\_famnrty\_famn\_for\_RTI\_GoKinD\_061118.xls - An Excel version of the SAS dataset. This data set is similar to the e1a\_famn family lookup table but each record has an additional flag if excluded from GAIN for any reason (ex\_gain); or because of site issues (ex\_site), quality control (ex\_qc), participant is parent in complete trio (ex\_parnt), or because of relatedness (ex\_relat).

### **3.2. GOKIND Reports Trios Probands**

Summaries of clinical and biochemical characteristics of eligible probands and the parents of eligible trio probands:

- ALLPARC2.PDF - Parents of eligible trio probands, all clinics
- ALLPRBC2.PDF - Eligible probands, all clinics
- GWUPARC2.PDF - Parents of eligible trio probands, GWU clinics
- GWUPRBC2.PDF - Eligible probands, GWU clinics
- JOSPARC2.PDF - Parents of eligible trio probands, Joslin Diabetes Clinic
- JOSPRBC2.PDF - Eligible probands, Joslin Diabetes Clinic.

### **3.3. GOKIND SAS to Excel**

The file a\_f\_ext1\_jul06\_v3\_20060729\_sas\_listing.doc describes “SAS library GK\_EXT1 Created for External Researchers 29July06,” which contains eleven members (data sets). These included data for Case Singleton & Complete Trios, Control Singleton and Complete Trios, and DNA Quality Control Samples.

The GOKIND\_SAS\_to\_Excel directory also contains the Excel versions of the SAS datasets that were created for external researchers. It contains two subdirectories, GOKIND\_SAS\_to\_Excel\_v1\_orig and GOKIND\_SAS\_to\_Excel\_v2\_formatted. The subdirectories contain files with the same names:

- e1a\_core\_GoKinD\_060729.xls - Core data (GW and Joslin), cases, controls and complete trio parents
- e1a\_famn\_GoKinD\_060729.xls - Family lookup at participant level
- e1a\_famr\_GoKinD\_060729.xls - Family lookup - 1 record per masked family
- e1a\_gene\_GoKinD\_060729.xls - Genetics for cases, controls and complete trio parents
- e1g\_core\_GoKinD\_060729.xls - GW cases, controls and complete trio parents
- e1g\_srbj\_GoKinD\_060729.xls - GW form BJFORM GK210
- e1g\_srbk\_GoKinD\_060729.xls - GW form BKFORM GK211
- e1g\_srb1\_GoKinD\_060729.xls - GW form BLFORM GK212
- e1g\_srbo\_GoKinD\_060729.xls - GW form BOFORM GK200
- e1j\_core\_GoKinD\_060729.xls - Core Data Joslin Participants
- e1j\_srce\_GoKinD\_060729.xls - Joslin Proband and Parent Medical Questionnaires
- ext1\_sd2\_contents\_GoKinD\_060729.xls - This spreadsheet summarizes the variable names for each data set, the column labels, the number of observations, and the number of positive observations.

### 3.4. GOKIND SAS Variables

This subdirectory contains variable descriptions and data dictionaries in a total of 17 files.

- Descriptions of SAS variables for the original SAS datasets in Excel format. This includes descriptions of SAS variables for internal use that are not included in the datasets released to external researchers. These descriptions are contained in Excel spreadsheets:
  - e1g\_core\_sas\_contents\_datadic\_060729.xls
  - e1a\_famn\_sas\_contents\_datadic\_060729.xls
  - e1a\_famr\_sas\_contents\_datadic\_060729.xls
  - e1a\_gene\_sas\_contents\_datadic\_060729.xls
  - e1g\_srbj\_sas\_contents\_datadic\_060729.xls
  - e1g\_srbk\_sas\_contents\_datadic\_060729.xls
  - e1g\_srbl\_sas\_contents\_datadic\_060729.xls
  - e1g\_srbo\_sas\_contents\_datadic\_060729.xls
  - e1j\_core\_sas\_contents\_datadic\_060729.xls
  - e1j\_srce\_sas\_contents\_datadict\_060729.xls
- Spreadsheet containing four worksheets with data dictionaries for all clinics, George Washington, Joslin, and the CDC genetic tests:
  - gk\_core\_data\_dic\_external\_060731.xls
- Study notes about specific patients:
  - Genetic sample quality - gk\_id\_pt1\_notes\_genetics\_20060706\_v19.xls
  - Families with sibs - gk\_id\_pt2\_notes\_families\_w\_sibs\_06Jul06.xls
  - Eligibility - gk\_id\_pt3\_notes\_elig\_20060706\_v19.xls
- Documentation of the SAS code used to create derived variables:
  - George Washington University - gokind\_deriv\_var\_gw\_v5\_20060509.doc
  - Joslin Diabetes Center - gokind\_deriv\_var\_joslin\_v5\_20060509.doc
- SAS formats:
  - sas\_formats\_gkfmtka2\_20060523\_v31.sas\_formats

### 3.5. GOKIND SAS Datasets

GOKIND sas7bdat files:

- e1a\_core.sas7bdat
- e1a\_famn.sas7bdat
- e1a\_famr.sas7bdat
- e1a\_gene.sas7bdat
- e1g\_core.sas7bdat
- e1g\_srbj.sas7bdat
- e1g\_srbk.sas7bdat
- e1g\_srbl.sas7bdat
- e1g\_srbo.sas7bdat
- e1j\_core.sas7bdat
- e1j\_srce.sas7bdat
- sas\_formats\_gkfmtka2\_20060523\_v31.sas\_formats.sas

**3.6. GOKIND Supplemental Data**

This folder contains a supplementary dataset from August 2008 that can be used to calculate the elapsed years between diabetes diagnosis and laser treatment without revealing the actual treatment dates. All dates were suppressed in the first data release in 2006 for confidentiality reasons. The folder contains the following three files:

- GoKinD\_supplement01\_data\_release\_20080801.doc – A description of the data files
- GoKinD\_retxpt\_20080801.sas7bdat - SAS data for the Diabetes Duration at Retinopathy data
- retxpt\_20080801.xls – Excel file for the Diabetes Duration at Retinopathy data

Form 220.2GK Clinic Forms Inventory (1 page)  
admin\_biform\_gk220v2\_20010406\_vars\_forms\_inventory.pdf

Form 227.3GK Lipid, Serum Creatinine, & Cystatin C Quality Control Mailing List (1 page)  
admin\_bqform\_gk227v3\_20020318\_vars\_mailqc\_lipid.pdf

Form 228.3GK Urine Quality Control Mailing List (1 page)  
admin\_brform\_gk228v3\_20020318\_vars\_mailqc\_urine.pdf

Form 229.3GK Genetics Quality Control Mailing List (1 page)  
admin\_bsform\_gk229v3\_20020318\_vars\_mailqc\_genet.pdf

Form 230.3GK Saved Specimen Mailing List (1 page)  
admin\_btform\_gk230v3\_20020318\_vars\_mail\_saved\_spec.pdf

Form 223.3GK Hemoglobin A1c Specimen Mailing List (1 page)  
admin\_buform\_gk223v3\_20020318\_vars\_mail\_hba1c.pdf

Form 222.3GK Genetic Specimens Mailing List (1 page)  
admin\_bvform\_gk222v3\_20020318\_vars\_mail\_genet.pdf

Form 224.3GK Lipid, Serum Creatinine, and Cystatin C Mailing List (1 page)  
admin\_bxform\_gk224v3\_20020318\_vars\_mail\_lipid.pdf

Form 225.3GK Urine Specimen Mailing List (1 page)  
admin\_byform\_gk225v3\_20020318\_vars\_mail\_urine.pdf

Form 226.3GK Hemoglobin A1c Quality Control Mailing List (1 page)  
admin\_bzform\_gk226v3\_20020318\_vars\_mailqc\_hba1c.pdf

Form 204.5GK Supply Reorder Form (1 page)  
admin\_gk204v5\_20021105\_coc\_supply\_reorder.pdf

Form 205.3GK CBL Supply Reorder Form (1 page)  
admin\_gk205v3\_20011022\_cbl\_supply\_reorder.pdf

Form 221.3GK Forms Mailing List (1 page)  
admin\_gk221v2\_20020708\_forms\_weekly\_cover.pdf

Form 202.2GK Personal Information on Proband/Relative (2 pages)  
gw\_bbform\_gk202v2\_20010406\_vars\_pers\_info.pdf

Form 203.2GK Notification of Death (2 pages)  
gw\_bcform\_gk203v2\_20010406\_vars\_death.pdf

Form 207.1GK Notification of Transfer, Remote Site Collection, or Refusal to Participate (2 pages)  
gw\_bgform\_gk207v1\_20020703\_vars\_transfer.pdf  
GOKIND Archive Description Version 2.0

Form 210.2GK Medical History and Physical Examination (8 pages)  
gw\_bjform\_gk210v2\_20010406\_vars\_med\_hx.pdf

Form 211.2GK Current Medication Form (4 pages)  
gw\_bkform\_gk211v2\_20010406\_vars\_curr\_meds.pdf

Form 212.1GK Historical Urine Value for Control Probands (1 page)  
gw\_blform\_gk212v1\_20020712\_vars\_historic\_urine.pdf

Form 200.3GK Checklist for Proband/Relatives (4 pages)  
gw\_boform\_gk200v3\_20020712\_vars\_checklist.pdf

Form 005a.10 GoKinD Study Diabetic Offspring (9 pages)  
joslin\_005av10\_20010329\_vars\_offspring.pdf

Form 005b.9 GoKinD Study  
Parents (7 pages) joslin\_005bv9\_20010302\_parent.pdf