Illumina Genotyping Control Database Purpose Document

Introduction

This document provides a description of the Illumina Genotyping Control Database (the "Database"). It also explains the processes and limitations of submitting data to or downloading data from the Database, and will help researchers better understand the purpose of the Database.

Background

Genetic case control association studies are becoming more widespread in the scientific community. As part of an ongoing initiative at Illumina, we provide access to genotypic and phenotypic data from individuals that can be used as controls in case control association studies, wherein risk factors of individuals with a certain disease (cases) are compared to individuals without the disease (controls). Illumina provides an online data repository that is accessible to the scientific community to further scientific progress using these resources, and is subject to appropriate terms and conditions.

Access and Data Security Overview

Access and interaction with the Database is restricted to individuals who are registered in iCom, Illumina's e-commerce system. All activity of Illumina-provided web services, including access to and use of the Database, are monitored by iCom login ID for usage patterns that are outside the policies of the respective resources. To access the Database, a user must have an active iCom account, and enter their secure iCom login ID and associated password.

To enroll a new iCom account

1. Go to https://icom.illumina.com/

2. Follow the online instructions for New Customer registration

3. An iCom login ID will be assigned

4. An email confirmation will be generated, and a reply is required to successfully complete the registration process

iCom login IDs are assigned a scope of resource access that is reviewed periodically by an internal committee responsible for maintaining the security and health of Illumina's web resources. iCom login IDs that have been inactive for six months are periodically verified using the email and organization information that is associated with each iCom login ID.

Workflow

Submitting Data to the Database

1. Login to the Database using iCom ID and password

2. Accept and adhere to the terms and conditions of the Illumina Genotype Control Database Submission Agreement (the "Submission Agreement").

3. Submit approval documentation confirming necessary approvals, consents, and authorizations to submit the data to the Database, as applicable to the genotypic and phenotypic data and as required by federal, state and institutional laws, regulations and/or policies. (See "Phenotype Data" and "Genotype Data" tables at the end of this document). Approval documentation may consist of local IRB approval confirming that an appropriate informed consent corresponds to the data to be submitted and authorizes the data to be submitted to the Database for the Designated Purposes, as well as a representative sample of the informed consent. The approval documentation is stored in digital files that once submitted, will be controlled under 21CFR part 11.

4. The submission event (submitter identifier, acceptance of the Submission Agreement, approval documentation, and genotype/phenotype data) is logged within a transactional database for tracking purposes.

    a. The genotype and phenotype data and the approval documentation are linked upon submission and uploaded into a staging area within the database that is accessible only to Illumina employees that are involved in the verification and validation of the data and the authorizations.

b. The genotype and phenotype data and approval documentation are reviewed/inspected while in the staging area.

c. The submitted genotype data is screened for completeness and correctness; samples will be eliminated if they exhibit a call rate of <95%.

d. Submitted data also undergo a quality control process where loci with poor call rates (<95%), or with a significant fraction of miscalled genotypes that result from poorly clustered data, are purged from the dataset.

5. Upon completion of the data inspection, the genotype data are released into the Database together with the phenotype data. Each sample maintains a database relationship with the respective approval documentation.

Searching and Downloading Data from the Database

1. Login to the Database using iCom ID and password.

2. Accept and adhere to the terms and conditions of the Illumina Genotype Control Database Download Agreement (the "Download Agreement"). Through the Download Agreement, researchers agree to be bound to using the data in the Database only for the Designated Purposes as outlined below. Acceptance of the Download Agreement is captured in a transactional database after which the researcher will have the ability to search and filter the database for samples based on phenotypic criteria.

3. Select a set of control data for download. The downloaded samples, along with the identity of the researcher, are captured in a transactional database supporting robust data insertion and rollback functions. All downloaded samples contain the phenotypic data associated with the criteria used during the search and filter process (?Where is the "Search and Filter" criteria described?)

Authorized Use of the Genotype Data (The "Designated Purposes")

Researchers downloading data from the Database must agree to be bound to use the data only for the following Designated Purposes:

1. The genotypic and phenotypic data can be used solely as part of an ongoing case control study. A case control study is defined as an epidemiological study in which the risk factors of people with a certain disease (cases) are compared with those without the disease (controls). The user can use the data from the Database solely for the purpose of statistical data analysis in combination with the genotypic data from the cases. The term "controls" in this context means individuals that have certain phenotypic characteristics, ethnic information, age, or have a specific gender that might be matched with the information for the individuals used as cases. If any of the controls have a non-matching phenotype, the user can exclude or include these individuals based on his Database search query.

2. The control genotype data can be used for software algorithm development. These algorithms should only focus on one of the following:

a. Addressing challenges associated with the analysis of sets of genotypic data.

b. Detecting differences in allele frequency based on phenotypic data.

c. Development of advanced analysis tools for the genetic research community.

The following are prohibited uses of the Database and the data contained therein:

1. Identification or attempts to identify individuals within the Database, including any downloaded data.

2. Use of the data from the Database to discriminate against any race, gender or any listed phenotypic data.

3. Use the data from the Database outside of the Designated Purposes.

4. Unauthorized distribution or sharing of data from the Database with other individuals.

5. Use of iCom login ID by an individual other than the individual corresponding to that iCom login ID.

No biological material or genotyping of new samples is provided as part of the Database program. Each individual submitting data to the Database must verify that appropriate approvals, consents, and authorizations were obtained to submit the data to the Database for the Designated Purposes, as required by federal, state and institutional laws, regulations and/or policies and in compliance with the Submission Agreement. Each individual downloading data must agree to be bound to use the downloaded data only for the Designated Purposes and in compliance with the Download Agreement. Illumina does not verify whether the approval documentation is sufficient to allow submission of data to the Database, and accepts no responsibility for any unauthorized use or disclosure of any data contained in the Database.

Referencing data from the Database in publications:

If any data from the Database are used in publication, use the following reference: "The control genotype data are from the Illumina Genotype Control Database (www.illumina.com) using the following search criteria (specify your search criteria and the date)."

Potential Future Application

Consistent with Illumina's intent to provide this Database to facilitate scientific progress, Illumina recognizes that this effort may be advanced by cooperating with others working on similar projects. Specifically, at some point in the future Illumina expects to formalize an arrangement to share data from the Database to dbGaP (the database of Genotype and Phenotype), a database designed to archive and make available data from genome wide association studies (GWAS). dbGaP was developed and is managed by the National Center for Biotechnology Information (the "NCBI"), which is a division of the National Library of Medicine, part of the National Institutes of Health.

This is a potential future application and not a current Designated Purpose. As such, for any dataset submitted to the Database before submission to dbGaP is identified as a Designated Purpose, Illumina will not provide any such dataset for submission to dbGaP without receiving approval from the submitter that such dataset can be submitted to dbGaP. Once submission to dbGaP is identified as a Designated Purpose, submitters can evaluate at the time of submission whether they can confirm that their dataset can be submitted to dbGaP.

Phenotype Minimum Information Required
Gender (M/F/U)
Ethnicity
Age (number, queried by range)*
Optional Phenotype Data
Weight (number – queried by range)
Height (number – queried by range)
Blood type
Blood type (+/-)
Product name and revision
Phenotype(s) positive
Phenotype(s) negative

*Ages >89 years are represented as "89 years and older"

Genotype Data Captured in the Database
Beadpool versions
Cluster file name
Beadpool manifest name
GenomeStudio version
Normalization transformations
Submitter of genotype data*
Representative informed consent documents*
Genotype Data Captured for each SNP
Genotype in the top strand ("--" if no call)
Genotype in forward strand ("--" if no call)
GenCall score
Raw intensities in x and y (GRN and RED)
Normalization transformation index
Genotype Data Provided for Downloading (for each SNP)
Genotype in the Top Strand ("--" if no call)
Genotype in Forward Strand ("--" if no call)

*Not provided to users

Illumina, Inc. • 9885 Towne Centre Drive, San Diego, CA 92121 USA • 1.800.809.4566 toll-free • 1.858.202.4566 tel • techsupport@illumina.com • illumina.com

FOR RESEARCH USE ONLY

© 2010 Illumina, Inc. All rights reserved.
Illumina, illumina*Dx*, Solexa, Making Sense Out of Life, Oligator, Sentrix, GoldenGate, GoldenGate Indexing, DASL, BeadArray, Array of Arrays, Infinium, BeadXpress, VeraCode, IntelliHyb, iSelect, CSPro, GenomeStudio, Genetic Energy, HiSeq, and HiScan are registered trademarks or trademarks of Illumina, Inc. All other brands and names contained herein are the property of their respective owners. Pub. No. 070-2007-004 Current as of 30 June 2010

illumina®

AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
TCAACGTACCGTAACGAACGTATCATTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTT
CGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACC
AACGTACCATTAAGAGCTACCGTGCAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAA
AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
GATTACTTGATCCACTGATTCAACGTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAGCAACGAGC
CGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGAAAAGAATGATAACAGTA