

文部科学省委託研究開発事業「統合データベースプロジェクト」
研究運営委員会(第7回)

ヒトゲノム多型・変異データ共有方針について

ライフサイエンス統合データベースセンター(DBCLS)
大学共同利用機関法人 情報・システム研究機構(ROIS)

箕輪 真理

2009年6月25日

統合DBプロジェクトでの共有方針検討の必要性

- 国内には疾患ゲノムの取り扱いに関する明文化されたルールがない！
⇒ 米国：NIH(dbGaP)、欧州：WTCCCなどはデータの共有・利用についてのルールを明文化されているのでそれらを参照。
- ゲノム情報は個人情報とつながる可能性も持つので、慎重な扱いが必要
⇒ 3省ゲノム指針、個人情報保護法も考慮する必要がある。どの情報を誰にどこまで公開できるか？必要な対策は何か？を実際の共有データについて検討。
- ゲノム特定での検討は共有方針までにとどまっている
⇒ 具体的な申請書や手順など、実運用に必要な書面の整備。

NIH(dbGaP)とWTCCCの状況

http://www.ncbi.nlm.nih.gov/gap

https://www.wtccc.org.uk/info/access_to_data_samples.shtml

The screenshot shows the dbGaP (Database for Genotypes and Phenotypes) website. The top navigation bar includes 'dbGaP GENOTYPES AND PHENOTYPES' and a search bar. Below the search bar, there are tabs for 'By Studies', 'By Diseases', and 'Advanced Search'. A table lists various studies with columns for Project, Study, Embargo Release, Details, Participants, and Type of Study. The 'Authorized Access' section is highlighted, providing information on how to apply for access, who can apply, and why access might be restricted.

Project	Study	Embargo Release	Details	Participants	Type of Study
CDR	AMD-MWAP Cohort Study: A Joint Genome Wide Association Study	Nov 28, 2009	[V][D][A]	3340	Case-Control
CDR	CDR: Genome Wide Association Study in Familial Parkinson Disease (FPO)	Feb 12, 2009	[V][D][A]	1991	Case-Control
CDR	CDR: Collaborative Study on the Genetics of Alcoholism (COGA)	Oct 06, 2009	[V][D][A]	1945	Case-Control
COO	Genome-Wide Association Study of Neuroblastoma	Dec 19, 2009	[V][D][A]	1032	Case-Control
GAN	Genotyping the 270 HapMap samples for GAN by Perlegen	-	[V][D][A]	-	Parent-Offspring Trios
GAN	Genotyping the 270 HapMap samples for GAN by Broad	-	[V][D][A]	-	Parent-Offspring Trios
GAN	Genome-Wide Association Study of Schizophrenia	Version 1: Nov 07, 2009 Version 2: Nov 07, 2009 Version 3: Aug 19, 2009	[V][D][A]	5066	Case-Control
GAN	Major Depression: Stage 1 Genomewide Association in Population-Based Samples	Version 1: Nov 07, 2009 Version 2: Nov 07, 2009	[V][D][A]	3741	Case-Control
GAN	Collaborative Association Study of Prostate	Aug 12, 2009	[V][D][A]	2875	Case-Control
GAN	Search for Susceptibility Genes for Diabetic Nephropathy in Type 1 Diabetes (GEMIN) study participants, GAD1	Apr 19, 2009	[V][D][A]	1825	Case-Control
GAN	Whole Genome Association Study of Bipolar Disorder	Version 1: Nov 23, 2009 Version 2: Dec 01, 2009 Version 3: Apr 20, 2009	[V][D][A]	3261	Case-Control
GAN	International Multi-Center ADHD Genetics Project	Nov 26, 2009	[V][D][A]	2935	Parent-Offspring Trios
GANZ	Genetic Multiple Sclerosis Association - GenesMGA	Nov 10, 2009	[V][D][A]	1920	Case-Control
GENEVA	A Genome Wide Scan of Lung Cancer and Smoking	Jan 08, 2010	[V][D][A]	5588	Case-Control

dbGaP Authorized Access
 dbGaP Authorized Access is the management portal for individual-level data. This site can be used to submit a data access request, manage access requests, and download approved data sets. [Log In to the Authorized Access system](#)

How does one apply?
 Click on the "log in" link (upper right side of this page) and follow the instructions. In order to log in and apply for authorized access to dbGaP studies you must have one of the following accounts:
 • eRA Commons (for NIH Extramural principal investigators, grantees, or other extramural investigators). If you do not have a pre-existing account, register here.
 • NIH Login (for intramural NIH scientists and staff)

For additional information, see [request procedures for Principal Investigators and Signing Officials](#).

Who can apply?
 Researchers from outside of NIH need to be identified as Principal Investigators (PIs) in the eRA Commons system. If you are not a PI, when you log in, the system will advise you to contact your local supervisor to apply on your behalf.
 NIH Intramural employees can apply if they are registered in the NIH Intramural Database (NIDB) as a PI, PI-eligible, or Lead Investigator. NIH Administrative and Extramural staff can apply if they meet specified criteria. When NIH employees log in, the system will recognize their status and direct them to obtain any necessary approvals and fill out any required forms. For further information, see NIH Feb. 6, 2008 memo on [Requesting Access to Data in the NIH GDSAS Data Repository](#).

Why is access restricted?
 NIH is committed to respecting the privacy and intentions of research participants with regard to how data pertaining to their individual information is used. Data access is therefore intended only for scientific investigators pursuing research questions that are consistent with the informed consent agreements provided by individual research participants. Furthermore, investigators provided access will be expected to utilize appropriate [PI data security measures](#).

Who is an authorized user within the data access request system?
 Authorized users include the researchers who may request data sets for specific research uses, the Institutional Signing Officials from the PI's home organization who certify and submit such requests, and the NIH staff who review and process requests (e.g., members of the Data Access Committee).

Do you need further help?
 dbGaP also maintains a [help desk](#) to assist investigators, institutional signing officials and NIH staff with authorized access management, and answer any questions related to the application process. Contact the [help desk](#) with your queries.

The screenshot shows the Wellcome Trust Case Control Consortium website. The main heading is 'Access to WTCCC genotype data'. The text explains the primary purpose of the WTCCC and provides information on how to access the data. It lists the types of data available and the conditions for access.

Access to WTCCC genotype data
 The primary purpose of the WTCCC is to accelerate efforts to identify genome sequence variants influencing major causes of human morbidity and mortality, through implementation and analysis of large-scale genome wide association studies. Additional objectives include the development and validation of informatics and analytical solutions appropriate to the scale and nature of the project, as well as use of the data generated to answer important methodological and biological questions relevant to association studies in general, and in the UK in particular (for example issues of population substructure).

The Consortium anticipates that data generated from the project will be used by others, such as required for developing new analytical methods, in understanding patterns of polymorphism and in guiding selection of markers to map genes involved in specific diseases.

Access to summary data and individual-level genotype data is available by application to the Wellcome Trust Case Control Consortium Data Access Committee. Access to data will be granted to qualified investigators for appropriate use. Individual-level genotype data and summary genotype statistics for WTCCC1 collections are held within the European Genome Archive. <http://www.ebi.ac.uk/ega>. For further information regarding EGA, please contact ega-admin@ebi.ac.uk.

Data available
 Summary statistics and individual-level genotype data is available for the following:

- Data from the following samples using the 500K Affymetrix chip:
 - 1,500 samples from the 1998 British Birth Cohort
 - 1,500 samples from the UK Blood Service Control Group
 - 2,000 samples each from the following disease collections: type 1 diabetes, type 2 diabetes, rheumatoid arthritis, inflammatory bowel disease, bipolar disorder, hypertension, coronary artery disease.
- Data from the following samples using the Affymetrix v6.0 chip:
 - 3,000 samples from the 1998 British Birth Cohort
 - 3,000 samples from the UK Blood Service Control Group
- Data from the following samples using a 1.5M Affymetrix chip:
 - 2,000 samples from the 1998 British Birth Cohort
 - 2,000 samples from the UK Blood Service Control Group

Two sets of genotypes are available for each of these, called by different algorithms: Chiamo as discussed and used in the analysis for the WTCCC papers and the Affymetrix algorithm BRLMM.

INFORMATION AND GUIDELINES FOR ACCESS TO DATA FROM THE WELLCOME TRUST CASE-CONTROL CONSORTIUM

The Wellcome Trust case-control Consortium Data Access Committee will consider applications for access to genotype data generated as a result of the Consortium activities. Access to data will be granted to qualified investigators for appropriate use. A qualified investigator means a scientific researcher who is employed, or a student enrolled at, or legitimately affiliated with an academic, non-profit or government institution, or a commercial company. The decision as to who is or is not a qualified investigator is the primary responsibility of the Consortium Data Access Committee.

Access is conditional upon availability of data and agreement to abide by policies related to publication, data/sample disposal, ethical approval and confidentiality.

Access to data will be governed by the provisions laid out in the associated informed consent and the original research ethics committee approval for each case collection or control group.

The Consortium Data Access Committee is concerned only with access to the core, anonymised, genotype data and samples generated by this study. The only phenotypic information held by the Consortium is that which is implied by membership of a particular case or control group. The Committee will not consider requests for more detailed phenotypic information that is held by the principal investigators for the individual case collections. Access to this data would be by arrangement with the relevant principal investigator.

The Committee will consider applications that include collaborators, but each Institution must sign a separate Data Access Agreement. Should you wish to share the data with additional collaborators not previously approved, they must make a separate application for access to the Data.

Members of the Consortium will have early access to genotype data relating to their own collections, to control data, and to preliminary analyses, but will be bound by the same agreements as external users concerning the use of these data.

策定までの経緯

1. 特定領域研究ゲノム4領域におけるGWASデータに関する検討(～2008.6)
2. ゲノム特定領域研究Medical Whole Genome Resequencing委員会におけるヒト全ゲノム配列解析研究についての検討(2008.7～)
3. 統合データベースプロジェクト・疾患解析DBグループ(代表機関、東京大学)におけるGWASデータ、Resequencingデータに関する検討(2008.11～)
 - 3-1. 疾患解析DB開発「倫理検討委員会」開催(第1回 6月5日)
 - 3-2. 同 上 (第2回 7月開催予定)

「倫理検討委員会」メンバー(敬称略)

位田隆一(委員長、京大)、井ノ上逸郎(東海大)、加藤和人(京大)、金森修(東大)、辻省次(東大)、福島義光(信州大)、武藤香織(東大)、米本昌平(東大)、徳永勝士(東大)

オブザーバー: 田中(文科省)、下川・井戸(医科歯科)、小池(日立製作所)、箕輪(DBCLS)

作成中の文書

- **統合データベースプロジェクト
ヒトゲノム(多型・変異*)データ共有方針**
公開されるデータの内容をカテゴリー・レベル分類し、それぞれについて、提供および利用の手続きについて記載。
- **申請書および報告書フォーマット**
データ提供申請書・データアクセス申請書・データ使用報告書。
- **データアクセス同意書**
データアクセスの条件等を記載。研究責任者の責務や事故時の対応など。
- **セキュリティ手順書**
データの安全確保のために実施すべき対策等を挙げたもの。

公開情報のカテゴリー（現在の案）

	レベル	1	2	3 “制限アクセスデータ”
格納データ	GWASデータ	A. 頻度データ、統計解析結果	B. 個体レベルでのCNV情報	C. GWAS遺伝子型データ D. GWAS生データ
	Resequencingデータ	E. リシークエンシングデータおよび変異データ（公知）	F. リシークエンシングデータおよび変異データ	（該当なし）
	付随臨床情報	（染色体上の位置、方向）・発症時あるいはサンプル収集時の年齢（10歳幅等）・臨床表現型に限った記述（病型及び病型に相当する情報など）・性別、といった個人を特定できない情報に限る		
手続	データ提供時	氏名、所属、住所（所属）、e-mailアドレス、データの概要などを所定のウェブサイトから入力し、データを「データ共有審査委員会」に送付する	（同左）	提供者が属する機関の倫理審査委員会およびLSDB「データ共有審査委員会」での審査が必要。データ提供申請書の提出。
	データ利用時	認証等必要なく閲覧可能（一般公開）	氏名、職名、連絡先、使用目的、e-mailアドレス（原則、所属機関から発行されたアドレス）を「データ共有審査委員会」に申請し、許可を受ける	提供者が属する機関の倫理審査委員会およびLSDB「データ共有審査委員会」での審査が必要。データアクセス申請書、データ使用報告書の提出。

制限アクセスデータの取り扱いに関する手続

