# AMP T2D Project

Data Transfer Agreement

THIS DATA TRANSFER AGREEMENT (the "Agreement') is made and entered into as of the ____ day of _____, 201_ (the "Effective Date"), by and between _____, a [State & company type], located at _____ ("Provider") and **The Broad Institute, Inc.,** a Massachusetts non-profit corporation, located at 415 Main Street, Cambridge, MA 02142 ("Recipient"). Collectively, Provider and Recipient shall be referred to as "Parties" and individually as a "Party".

**WHEREAS,** Accelerating Medicines Partnership ("AMP") is a public-private partnership between the National Institutes of Health ("NIH"), the U.S. Food and Drug Administration ("FDA"), 10 biopharmaceutical companies and multiple non-profit organizations aiming to transform the current model for developing new diagnostics and treatments by jointly identifying and validating promising biological targets for therapeutics with the ultimate goal of increasing the number of new diagnostics and therapies for patients and reducing the time and cost of developing them; and

**WHEREAS,** Recipient will serve as the Data Coordinating Center ("DCC") of AMP in Type 2 Diabetes ("AMP T2D") and on behalf of AMP T2D, the DCC will aggregate data, automate analysis, and communicate results relevant to the genetics of type 2 diabetes and related traits, and support collaboration within the AMP T2D project ("AMP T2D Project"); and

**WHEREAS,** Provider has agreed to provide certain de-identified human genetic data ("Data") to Recipient in order to further the mission of the AMP T2D Project; and

**WHEREAS,** Recipient will make the data derived from genetic analyses of the Data ("Aggregate Data") available through a web-based portal ("AMP T2D Portal") to other members of the AMP T2D Project as well as share with the broader biomedical research community (the "Purpose");

NOW THEREFORE, it is mutually agreed as follows:

1. Provider agrees to transfer the Data described in Exhibit A for use in the AMP T2D Project.

2. Provider represents that the Data will be de-identified and all Protected Health Information, as defined by the Federal Health Insurance Portability and Accountability Act (HIPAA, 45 C.F.R. 164) ("HIPAA") will have been removed prior to sending the Data to Recipient. Notwithstanding the foregoing, if Recipient believes it has received identifiable patient information hereunder, it will hold such information in strict confidence indefinitely, immediately inform Provider and comply with Provider's instruction at Provider's expense, with respect to return or destruction of the same.

3. Recipient agrees to retain control over Data received from Provider and further agrees not to provide the Data, with or without charge, to any other entity or any individual other than Principal Investigator. Notwithstanding the foregoing, Provider acknowledges and agrees that Recipient may

provide the Data to the University of Michigan for use in its protected imputation server in furtherance of the AMP T2D Project and in accordance with Exhibit C. Nothing in this Agreement shall preclude the Provider from transferring its solely owned Data to other third parties for commercial or research purposes.

4. Recipient shall use reasonable technical, administrative, and physical safeguards to protect the Data from unauthorized access, use, or transmission. Recipient shall promptly (in no event less than one (1) business day) notify Provider if any Data is accessed, used, or transmitted in an unauthorized manner.

5. Recipient shall use appropriate technical and organizational security measures to protect the Data against unauthorized or unlawful processing and against accidental loss, destruction, damage, alteration or disclosure, by (i) ensuring a level of security appropriate to the harm that may result from such unauthorized or unlawful processing or accidental loss, destruction, damage, alteration or disclosure, and appropriate to the nature of such Data, (ii) regularly testing such measures to validate appropriateness and effectiveness, and (iii) implementing corrective action where deficiencies are revealed by such testing.

6. Recipient shall comply with all applicable laws, rules, and regulations related to the use of such Data, including but not limited to HIPAA.

7. This Agreement is not transferable to another facility that is not under the control of Recipient.

8. No option, license, or conveyance of rights, express or implied, is granted by Provider to Recipient in connection with the Data provided under this Agreement, except the right to use the Data strictly in accordance with the terms of this Agreement.

9. No rights of Recipient under this Agreement may be assigned or otherwise conveyed to any party, including a purchaser of Recipient, without the specific written agreement of Provider.

10. Provider acknowledges that Recipient shall make the Aggregate Data available to users through the AMP T2D Portal. Recipient agrees that Provider shall have access to the Aggregate Data through the AMP T2D Portal, at no charge, for its research, education and publication purposes.

11. Recipient agrees to implement a policy, as further detailed in Exhibit B, which will require users of the AMP T2D Portal to appropriately cite data acquired from the AMP T2D Portal in their publications. Notwithstanding the foregoing, in the event of any inconsistency between this Agreement and Exhibit B, this Agreement shall control.

12. Recipient agrees that it will not use the Data or Aggregate Data to establish the individual identities of any of the subjects from whom Data and Aggregate Data were obtained.

13. NO WARRANTIES, EXPRESS OR IMPLIED, ARE OFFERED AS TO THE MERCHANT ABILITY OR FITNESS FOR ANY PURPOSE OF THE DATA AND AGGREGATE DATA PROVIDED UNDER THIS AGREEMENT. THERE ARE NO WARRANTIES OR REPRESENTATIONS AS TO THE PURITY, ACCURACY, SAFETY OR USEFULNESS OF

THE DATA OR THAT THE USE OF THE DATA WILL NOT INFRINGE ANY PATENT OR OTHER PROPRIETARY RIGHT. Unless prohibited by law, Recipient assumes all liability for claims for damages against it by third parties arising from the Recipient's use of the Data except when such claims are caused by the gross negligence or willful misconduct of Provider.

14. Provider certifies and Recipient acknowledges that the conditions for use of Data and Aggregate Data are approved by the Institutional Review Board (IRB) of the Provider in accordance with Department of Health and Human Services regulations at 45 CFR Part 46. Recipient agrees to comply fully with all such conditions with respect to making the Aggregate Data available to members of the AMP T2D Project as well as the broader scientific community. Recipient remains subject to applicable laws or regulations and institutional policies which provide additional protections for human subjects.

15. Recipient agrees to use the Data solely for the Purpose and the terms of this Agreement.

16. Provider may terminate this Agreement if Recipient is in breach of this Agreement and if such breach has not been cured within thirty (30) days after the date of written notice by Provider of such breach. Upon termination of this Agreement, Recipient agrees to return all Data to Provider, or provide Provider with written certification of destruction of the Data, at the election of Provider.

17. Failure to comply with any of the terms specified herein may result in disqualification of Recipient from receiving additional Data from Provider.

18. Recipient acknowledges that damages resulting from a breach of this Agreement may be difficult, if not impossible, to measure accurately, and the injuries sustained by the Provider in the event of a breach may be incalculable and irremediable. Therefore, the Provider may be entitled to seek an injunction, specific performance or other equitable relief in the event of a breach or threatened breach by the Recipient of their obligations hereunder and such right shall be cumulative and in addition to any other remedies which may be available at law or in equity.

19. The parties will comply with all applicable international, national, state, regional and local laws and regulations, including all applicable import and export control laws, in exercising their rights or performing their duties under this Agreement.

20. Unless otherwise specified, this Agreement embodies the entire understanding between Provider and Recipient with respect to the subject matter hereof, and any prior or contemporaneous representations thereof, either oral or written, are hereby superseded. Amendments to this Data Transfer Agreement must be made in writing and agreed to by both Parties.

21. This Agreement and any amendment hereto may be executed in counterparts and all such counterparts taken together shall be deemed to constitute one and the same instrument. If this Agreement is executed in counterparts, no signatory hereto will be bound until all the Parties named below have duly executed a counterpart of this Agreement.

IN WITNESS WHEREOF, the Parties hereto have duly executed this Agreement as of the Effective Date by their authorized representatives.

_____

By: _____

Name: _____

Title: _____

**THE BROAD INSTITUTE, INC.**

By: _____

Name: _____

Title: _____

**Exhibit A**

**Data Description Template**

To be completed by submitter with signed DTA for Study intake summary information.

| | |
|---|---|
| Study Name: | |
| PI(s) | |
| PI email(s) | |
| Submitter name(s) | |
| Submitter email(s) | |
| Type of data to be submitted | |
| Total # of samples to be submitted | |
| Project name | |
| Project description | |

## Exhibit B

Policies and Plan for Data Submitted to the DCC and Shared on the T2D Knowledge Portal

The Broad Institute will serve as the Data Coordinating Center (DCC) for AMP-T2D. On behalf of AMP-T2D, the DCC will aggregate data, support analyses, and continue to update capabilities to disseminate results relevant to the genetics of type 2 diabetes (T2D) and related traits, while coordinating collaboration within the AMP-T2D Project. The DCC will also be responsible for sharing the results from the data coordination and analysis activities on the AMP-T2D Knowledge Portal (T2DKP).

**Data Aggregation, Analysis, and Resource Distribution**

As the DCC for the AMP-T2D Consortium, the Broad Institute (along with its AMP funded partners) intends to (a) serve as the gateway to a large (and growing) aggregation of data relevant to the genetics of type 2 diabetes and its complications; (b) perform and automate analyses required to interpret those data; and (c) communicate results to diverse audiences via an open access Web Portal (T2DKP), presentation and publication. Each of these goals involves distinct categories of resources and activities that we will share and/or manage:

***Data aggregation***

We will aggregate data on behalf of AMP-T2D. Data aggregated under this effort will not be generated by the Knowledge Portal development work funded through the Portal-specific grants, but rather obtained from other investigators and repositories who wish to collaborate on and contribute to this effort, including other investigators funded by the wider AMP-T2D program. Moreover, the role of the Portal will not be to redistribute individual-level data, but rather to generate results attained via standard and customized queries that can be widely shared with the scientific community. Because the primary, individual-level data are neither generated by this project, nor redistributed to other users, the role of the Portal is limited to secure intake, storage and management, automated analyses, and dissemination of results in summary (*i.e.*, not individual-level) form while complying with intended use of the data and all relevant regulations.

We will focus on three classes of data: individual-level genotypes, individual-level phenotypes, and external precomputed results or annotations (*e.g.*, results from individual studies or meta-analyses of multiple studies, processed annotations).

The data and results currently stored in the Portal, as of September 2016, have either been generated at the Broad Institute as part of IRB-approved secondary use protocols or, in the case of meta-analysis results from published GWAS datasets, obtained in such a form that Broad was determined to be "not engaged in human subjects research" (per the criteria described in the U.S. Health and Human Services Office for Human Research Protections' 2008 Guidance on Engagement of Institutions in Human Subjects Research, under criterion #7; see also http://www.hhs.gov/ohrp/policy/cdebiol.pdf). All data have been de-identified prior to being sent to Broad; at no time or under no circumstances will investigators funded by this grant have information linking data back to subject identifiers.

For datasets subsequently added to the Portal, raw data will be obtained through formal NIH systems for data-sharing such as dbGaP, or directly from investigators who collected the data. It will be the decision of NIH and the AMP-T2D Steering Committee whether to accept into the Portal data that are not in dbGaP and, if so, the terms on which the data can be made accessible for analyses by other parties. For the raw data transferred to the DCC for representation on the Knowledge Portal, in all cases, we will obtain a DTA between the Submitter and the Broad Institute (outlined below). The DTA outlines the use and protection of the transferred data. The Submitting site will be responsible for ensuring that the datasets transferred to the Broad are consented for transfer, genetic analysis, and representation on the T2D Knowledge Portal. At the DCC, we will develop software for storing and managing datasets, and will not redistribute the raw data to any outside third parties. However, as part of our analysis process, we may send array data to the University of Michigan (an AMP-T2D funded site) imputation server for imputation purposes only. See Appendix C of our data transfer agreement for additional details. For bulk data (*e.g.*, raw and harmonized individual-level genotype data), we will use object storage systems, with access controlled through application programming interfaces (APIs). Only authenticated and authorized users can access data; all such access is logged and auditable.

Additionally, we are currently building a database tool that captures data use restrictions for each dataset electronically using an ontology-based consent database, and can match those restrictions against potential research usage to ensure that only appropriate users can query specific datasets. Web-based tools, currently in development, support both the entry of data-use restrictions and the review of access requests. This will enable the aggregation of additional and more diverse datasets for the Portal.

***Examples of datasets:***
1. Current and future genetic studies of type 2 diabetes
2. Current and future genetic studies of related quantitative traits
3. Current/future studies of type 2 diabetes-related complications
4. Annotations of function

***Classes of datasets for storage and analysis in the Portal***

1. Those that ***do not require*** ethical and regulatory approval
    - Results from publicly available datasets
    - Summary statistics
2. Those that ***require*** approval from public access sites
    - Access controlled export (dbGaP, EGA, etc.) of individual-level data where we will serve results and summary statistics
3. Those that ***require*** local IRB approval

- Data generated directly at the Broad on de-identified DNA samples

4. Those that ***require*** a Data Transfer Agreement (DTA)
    - All de-identified individual-level genetic and phenotypic data generated externally

***Data transfer agreement***

For all de-identified individual-level genetic and phenotypic data generated externally (genotype, phenotypes, annotations, etc.) received by the Broad Institute as DCC for the AMP-T2D Knowledge Portal, we will execute a Data Transfer Agreement (DTA) with the submitting institution. We will ensure that the usage of the data is compliant with the Data Use Restrictions associated with the dataset. It will be the responsibility of the submitting institution to outline the Data Use Restrictions for the data coming to the DCC for the Knowledge Portal. It will be the responsibility of the submitting site or Institution to outline the appropriate Data Use Restrictions as part of the executed DTA for the DCC. Below we outline the data use and analysis plan for the DCC.

***Data use and analysis***

For all de-identified individual-level genetic and phenotypic data generated externally (genotype, phenotypes, annotations, etc.) received by the Broad Institute as DCC for the AMP-T2D Knowledge Portal we will perform quality control assessment, harmonization, and association analysis for T2D and related traits. This process will be a collaborative "handshake" process with the submitter. At the completion of each major phase we will share a report and results with the submitter. All results will be approved by the submitting site before results are available through the T2DKP. We have outlined our procedures for the data use, and analysis steps in a document entitled "AMP T2D Portal Submitter and Analysis Guide for Data at the DCC," which may be downloaded at: https://s3.amazonaws.com/broad-portal-resources/AMP_T2DKP_Submitter_Guide.pdf

We request a number of traits for association analysis at the DCC for the purposes of deposition in the AMP-T2D Knowledge Portal, including: Type 2 Diabetes, Fasting Insulin, Fasting Glucose, Lipids (HDL, LDL, TG, Total Cholesterol), Blood pressure, Creatinine, BMI (height and weight), waist circumference, lipid medications. We encourage submitters to share additional phenotypes, where approved by their local IRB/Ethics Committee (based on the cohort-level patient consent forms), as outlined in the AMP T2D Submitter's Guide to Sending Data to the DCC.

Assuming the submitting site has approved all these traits for analysis and display on the T2DKP, we will perform association analysis in the following tiered manner:

- Phase 1: T2D status, fasting glucose, fasting insulin

- Phase 2: 2hr glucose, 2hr insulin, HbA1C, HOMA-B, HOMA-IR
- Phase 3: BMI, HDL, LDL, triglycerides, total cholesterol, diastolic blood pressure, systolic blood pressure, WHR, waist circumference, hip circumference, height
- Phase 4: Longitudinal and Complications data: specification for these trait types have yet to be defined by the AMP-T2D Phenotype Working Group

***Results Sharing:***

The DCC will share the results of all approved analyses directly with the submitter upon completion for review. We will partner to address any quality control matters or confounders in the data before deposition in the T2DKP. Once the results are finalized, the DCC will make the data available for query via the T2DKP.

The individual-level data sent by data submitters, stored, and analyzed by the DCC will never be shared with Portal users; only results will be shared. The individual-level and summary data will reside in one or several data vaults behind a secure firewall. User-activated analytical modules will be deployed behind the firewall to analyze the data or query precomputed results. The Portal will provide results in response to queries for information, obtained from genetic analyses performed on the data. The purpose of the AMP-T2D web Portal will be to enable broad access to the comprehensive results of genetic studies of type 2 diabetes, related traits, and diabetic complications. To ensure that the web Portal is effective in allowing access to results and data – both within AMP-T2D and with the broader biomedical research community – we will develop an interface to provide access to results in a form designed to meet user needs while maintaining the individual data privacy requirements, and will engage Portal users in assessing the value of these features.

Results from studies included in the Portal will be available genome-wide (*i.e.*, not limited to "top hits"), and results from different studies and types will be integrated and presented simultaneously. In most cases the results on the T2DKP will be queryable by study. For studies representing multiple cohorts with differential data sharing approvals, we will display the approved results by cohort. Where a study represents multiple ethnicities, we will also allow query of the results by cohort-reported ethnicity. Metadata and other technical details (*e.g.*, analysis parameters, explanations of terms, documentation of methods) will be available at lower levels of drilldown.

**Resource Distribution and Sharing**

We will share software, methods, and code developed as part of consortium efforts. Specifically, we envision three types of sharing: (a) sharing of software source code; (b) sharing services; and (c) sharing of effort between groups with the intention of maintaining or extending existing software.

***Sharing of software source* code**

We are producing open-source software under the terms of the BSD 3 open-source license (see http://opensource.org/licenses/BSD-3-Clause). As such, this code will be freely available for use by any other parties; the software will be supplied "AS IS" with no implied warranty or promises of support. We will maintain a Github repository from which interested parties can download the source code. The code source of the AMP-T2D Portal, entitled, Framework for investigating Genetic Associations (FGA) is located here: https://github.com/broadinstitute/dig-diabetes-portal.

***Sharing services***

Our software will be constructed as a distributed system in which computers communicate using standard protocols (HTTP for transport, with REST as an organizing principle and data payloads defined with JSON), with well-defined interfaces specific to the computational topics addressed by each computer system. These services will in principle be accessible by any other party willing to adopt the conventions used by our services. To the extent that data on these services may be under privacy and use restrictions, these running services will be designed to provide information only in forms that protect privacy and security, or in secure, encrypted mode for other parties with permission to access and receive the information.

***Sharing of effort between groups to maintain or extend existing software***

The Portal architecture will be designed to facilitate front-end contributions (*e.g.*, extensions of existing widgets for data exploration) from a wide community of developers. Data or computations for REST servers will be encapsulated as loosely coupled "plug-in" modules that may be written in different languages (*e.g.*, Python, JVM-based languages, shell scripting). This approach anticipates the contribution of computational modules from other individuals and groups, both within and outside of the AMP consortium.

**Policies for Data Release, Accessing the Portal and Terms of Conduct**

***Data processing and availability (applicable for both data coming to DCC and to Federated nodes)***

We will have a 3-stage process for data release on the Portal. The figure below outlines the stages and timelines. The stages are:

1. Data Deposit: The DCC (and Federated nodes) will receive data from submitters on an ongoing basis. The Data deposition stage has several components that must be completed for the data to be ready for release into the Portal. These are:

(1) Data use agreements and ethical approvals for data transfer to DCC (or the Federated nodes) and release into Portal.
(2) Physical transfer of data and all meta-data, in required formats, into Data Intake System at the DCC (or at a Federated node).
(3) Data storage, curation, QC, and harmonization.

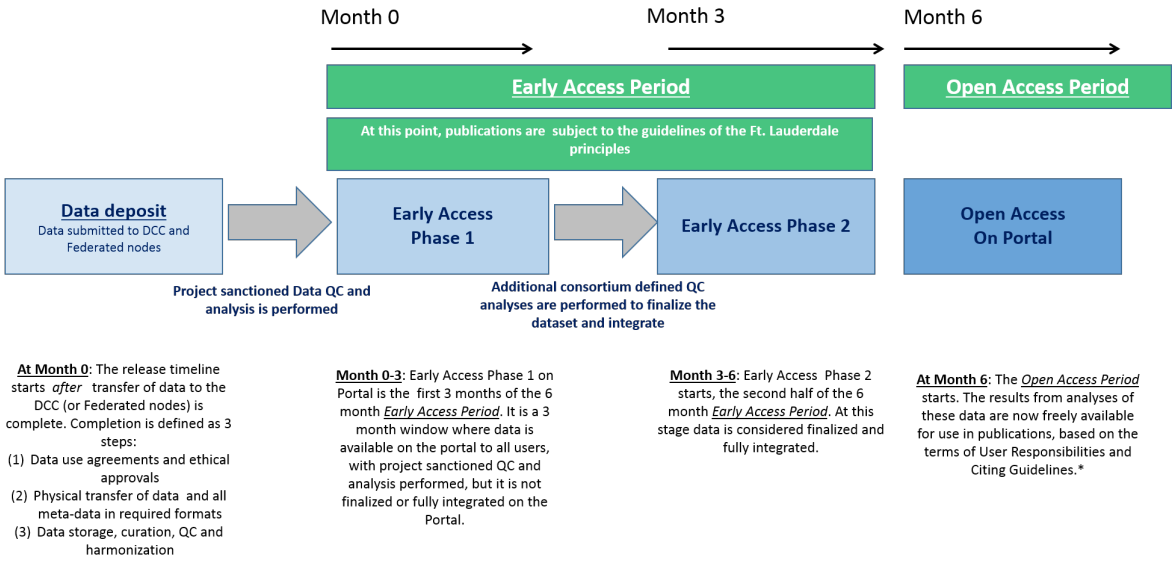
In general, upon depositing data into the Portal, a QC filter on genotypes and phenotypes will be deployed as per standard operating procedures in the field. Data will only be available after these initial filters are applied. Filters will include automated steps and final human curation, as determined by the AMP-T2D investigative team.

2. Early Access Period: After the Data deposition steps are complete, data will be released to the Portal. This denotes the start of the Early Access Period. This 6-month time period will be divided into 2 phases. The first 3 months will be the "Beta release" window where data are available on the portal to all users. At this point, project sanctioned QC and analyses have been performed, but the data are *NOT* considered final or fully integrated on the Portal. The goal of the "Beta release" time period is to allow users to review the data, to perform additional QC and analyses (ideally in a 'crowd-sourced' manner), and to finalize and integrate the dataset. At the end of the first 3 months, the data will move to the "Alpha release" state, where data are considered final and fully integrated.

During this 6-month period, all analyses, results, and publications proposed are subject to the "Fort Lauderdale Principles" articulated for the sharing of genomic data (https://www.genome.gov/pages/research/wellcomereport0303.pdf).

3. Open Access Period: After 6 months, the Early Access Period ends, and there is Open Access to the finalized and fully integrated datasets. The results from analyses on the data are now *freely available* for use in publications, based on the terms of User Responsibilities and Citing Guidelines.

***Data release timeline:***




***Data use and availability***

All users are welcome to use results from analyses of the data in the Portal to further their research without seeking explicit permission from the Portal team or funders. Users are also welcome to cite data in scientific publications, provided that they cite the Portal as the source. If users are citing a single dataset represented in the Portal, they should cite both the Portal and the relevant paper for that dataset (if one has been published).

The following polices are listed on the Portal and each user who logs on will be required to read these terms and click an acknowledgement of conduct before accessing the Portal:

***User registration***

To access the Portal, users must obtain a Google ID, which will be used for quality control (QC) and monitoring purposes (see below). In the future, should the AMP-T2D Consortium and Data Submitters agree, we may develop a more stringent registration process, requiring identification and authentication of the user and institutional affiliation.

**User Responsibilities**

Portal users are expected to abide by the following provisions on data use:

1. Users will not attempt to download any dataset in bulk from the Portal
2. Users will not attempt to identify or contact research participants
3. Users will protect data confidentiality
4. Users will not share any of the data with unauthorized users
5. Users will report any inadvertent data release, security breach, or other data management incidents of which they become aware
6. Users will abide by all applicable laws and regulations for handling genomic data
7. Users will not submit a manuscript for publication until the Early Access Period is over (6 months after the clean dataset becomes available in the Portal), to allow for beta testing on the integrity of the dataset and finalization of the results on the Portal.

Agreeing to these provisions is a requirement of Portal use. Violating them may result in an NIH investigation and sanctions including revocation of access to the Portal.

***Citing portal data***

Users who wish to cite data in this Portal in a scientific publication should do so in the following format:

AMP-T2D Program; T2D-GENES Consortium, SIGMA T2D Consortium. Year Month Date of access; URL of page you are citing.

For instance, a user who viewed the Portal's page on the gene *SLC30A8* on February 1, 2015, and wanted to cite it would use this citation:

AMP-T2D Program; T2D-GENES Consortium, SIGMA T2D Consortium. *SLC30A8*. type2diabetesgenetics.org. 2015 Feb 1; http://www.type2diabetesgenetics.org/gene/geneInfo/SLC30A8.

The Portal does not yet have a PubMed identifier.

***Re-using written content on the portal***



***User tracking***

The Portal team tracks a limited set of usage statistics. We do this to improve functionality based on how users interact with the Portal and to ensure that Portal data are being used properly (see our data use policy). Two types of people are allowed to view usage statistics at different levels of detail:

- Our website developer tracks de-identified, aggregate analytics (such as hit counts for specific pages) in order to improve the Portal's user experience. He does not

view statistics attached to individual user accounts.

- NIH personnel may be asked to examine individual user histories in cases of suspected misuse of Portal data.

## Exhibit C

## Secure Transfer of data to University of Michigan Imputation Server

.

As part of the analysis process at the DCC, for genotype and array data, we may run imputation on the data before deposition into the Knowledge Portal. The University of Michigan is our collaborative and funded partner for the AMP T2D KP effort and has developed a secure cloud-based imputation server: (https://imputationserver.sph.umich.edu/index.html) to impute genetic data against the most complete reference panel published to date. The server offers imputation from HapMap, 1000 Genomes (Phase 1 and 3), CAAPA and the updated Haplotype Reference Consortium (HRC version r1.1) panel. A complete description of how data are processed and protected during imputation is supplied on the Imputation Server site. The DCC will securely transfer the data to the Imputation Server solely for the purposes of imputation. The results will be retrieved directly by the DCC for additional analyses. Once the data are downloaded from the Server, they are deleted.