1.0 PURPOSE

1.1 The purpose of this standard operating procedure (SOP) is to outline the process and content of data exported from the Genotype-Tissue Expression (GTEx) Project to all external receiving entities.

2.0 SCOPE

2.1 The SOP covers the procedure for providing and controlling the content of data exports, the process for data exports to entities, and procedural verification for appropriate receipt, review, and security at the receiving entities.

2.2 Receiving entities for this project include: the Brain Bank and the Laboratory, Data Analysis and Coordinating Center (LDACC).

2.3 Currently all data exports outside of the Comprehensive Data Resource (CDR) are de-identified.

2.4 Export to additional entities would require review and approval by the GTEx project management team.

3.0 RESPONSIBILITY

3.1 **Data management team** is responsible for the dissemination of clean data to receiving entities and to identify incomplete, incorrect, inaccurate, irrelevant, etc. data, and validate or correct the values by replacing, modifying or deleting inconsistencies in the CDR and sources of data for the CDR.

3.2 **Project Team** is responsible for reviewing incoming data requests with all appropriate parties for approval for release to receiving entities. This also includes review and approval for the methods for release of the data, format, and content of data release and frequency of release. The GTEx project team is also responsible for ensuring that receiving entities are made aware of what their roles and responsibilities are in the receipt of the data and ensure that they have SOPs in place for receipt, review, and security of the data exported.

3.3 **Provider** includes all entities participating in the review and dissemination of the data export. This includes the CDR, the data management team, and the GTEx project team.

3.4 **Receiving entities** shall have SOPs in place for:

3.4.1 Identifying requested data elements for export. This includes a change to any previous requested data set template.

3.4.2 Confirming receipt of data export using email or other approved means.

3.4.3 Identifying authorized personnel qualified to receive the export.

3.4.4 Identifying and notifying the data management team with any questions or discrepancies noted upon receipt of the data export. This includes receipt of data elements not approved for receipt.

3.4.5 Confirming adequate storage and security of the received data export, including a security plan minimizing risk of unintended access or use.

4.0 DEFINITIONS & ACRONYMS

4.1 **Brain Bank** – Enlisted by the National Institutes of Mental Health (NIMH) for the as the receiving and processing facility for the brain, brain stem, and hair biospecimens collected for the GTEx project.

4.2 **BSS** – Biospecimen Source Site. Hospitals and/or research facilities tasked to collect, process, store, and ship clinically-annotated biospecimens and associated data in accordance with program-developed SOPs and protocols.

4.3 **CBR -** Comprehensive Biospecimen Repository. Centralized entity responsible for creating and managing the kit components shipped to the source sites. Also responsible for informatics support with the CDR, storage and processing all designated biospecimens (and associated inventory data) collected, creation of slides and digital images, and maintenance of the image management system, as well as quality checks on collected and processed specimens.

4.4 **CDR** - Centralized custom-made informatics system that stores and reports all collection, handling, and processing data for biospecimens and annotations collected for use by this program. The system provides secure, role-based access for BSSs to input data related to each case collected that is associated with a Limited Data Set (LDS) related to the donor. Interfaces are provided to other systems that contain related case data (e.g., inventory data at the CBR, molecular data at the molecular analysis facility, research data in dbGaP at the LDACC, etc.).

4.5 **Clean Data** – Data associated with a Case ID that is been reviewed for errors characterized by contradictions, disparities, keying mistakes, missing information, etc. Data is confirmed clean when these errors have been identified and corrected by the data management and /or project management team following the appropriate process

4.6 **Data Elements** - Refers to all operational, clinical, and pathology review data or information, related to a case collected for GTEx. This may include collection, handling, shipping, medical, clinical, pathological and social history data points that have been collected at any time point during the project.

4.7 **Data Export Template** - The data export template is a document that defines the data elements appropriate for the receiving entity to receive per the MTA/DUA (see definition below) and project specific policies. The template is

approved by the Project Team and will serve as the standard export for the CDR. This template will be recipient specific, and revised accordingly to suit the needs of the recipients and the project. The template will be version controlled.

4.8 **Data export** – The formatting and transfer of de-identified data that is used by another application, enabling the two programs to electronically access the same information.

4.9 **Data Export Workflow –** Encompasses all touch points from data abstraction of approved data elements to confirmation of receipt at the intended recipient/receiving entity.

4.10 **De-identified data** – Data which has been stripped of the 18 HIPAA (see definition below) identifiers and for which there is no knowledge that the remaining information could be used to identify the individual (see HIPAA Privacy Rule).

4.11 **GTEx** – Genotype-Tissue Expression. An NIH Common Fund program initiated to determine how genetic variation controls gene expression and its relationship to disease. GTEx aims to study human gene expression and regulation in multiple tissues, providing valuable insights into the mechanisms of gene regulation and, in the future, its disease-related perturbations. Genetic variation between individuals will be examined for correlation with differences in gene expression level to identify regions of the genome that influence whether and how much a gene is expressed. GTEx provides a base set of biospecimens and data whose quality is verified by the Pathology Resource Center (PRC) prior to being analyzed by the LDACC.

4.12 **HIPAA** – Health Insurance Portability and Accountability Act. Statutory law that governs the use and disclosure of Protected Health Information. Commonly known as the HIPAA Privacy Rule, this law was created to protect the privacy of health information that identifies an individual while still allowing other activities of benefit to society, such as research. While the HIPAA Privacy Rule does not apply to biospecimens directly, it may affect biospecimen resources that are considered covered entities in that human specimens often are accompanied by identifiable protected health information

4.13 **LDACC** – The Laboratory, Data Analysis and Coordinating Center. Research institute responsible for the overall coordination of GTEx activities and molecular and statistical analysis laboratory.

4.14 **LDS** – Limited Data Set. An exception to the HIPAA Privacy Rule requirement for an authorization from the subject for research use of protected health information. Refers to PHI (see definition below) that excludes 16 categories of direct identifiers and may be used or disclosed, for purposes of research, public health, or health care operations, without obtaining either an individual's authorization or a waiver or an alteration of authorization for its use and disclosure, with a data use agreement.

4.15 **MTA/DUA** – Material Transfer/Data Use Agreement. An agreement that governs the transfer of tangible research materials and data between two organizations, when the recipient intends to use it for his or her own research purposes. It defines the rights and obligations of the provider and the recipient with respect to the use of the materials.

4.16 **PHI** – Protected Health Information. Individually identifiable health information as defined in the HIPAA regulations (45 CFR § 160.103), that is or has been collected or maintained by the covered entity in the course of providing healthcare that can be linked back to the individual participant.

4.17 **PRC** – Pathology Resource Center. Centralized resource consisting primarily of expert pathologists whose function is to review biospecimens (either slides or images) collected for the program and assess their quality and fitness for use by researchers.

4.18 **Provider** – Entity or person(s) that is/are providing the materials and information on behalf of the project.

4.19 **Recipient –** The entity that receives the export from the Provider (i.e. the LDACC, Brain Bank).

4.20 **SOP** – Standard Operating Procedure. An established procedure to be followed in carrying out a given operation or in a given situation.

4.21 **CBR IS** - CBR Informatics System - Advanced biospecimen inventory management system provided and administered by the Comprehensive Biospecimen Resource (CBR) designed to track the lifecycle of biospecimens, as defined by the projects' objectives from the point of collection through delivery, receipt, storage, processing and distribution to various receiving entities.

5.0 ENVIRONMENTAL HEALTH & SAFETY

5.1 Not applicable

6.0 MATERIALS/EQUIPMENT

6.1 Not applicable

7.0 PROCEDURE FOR PROVIDER

7.1 GTEx project team will request verification from the receiving entities that they have locally established SOPs for receipt, review and security of the data they receive, include all of the following elements, and be verified by the project team for acceptability:

7.1.1 Authorized personnel to receive the data export.

7.1.2 Any data export discrepancies shall be communicated and resolved through the joint efforts of the GTEx DM and informatics teams and the

receiving entity's staff. This includes data elements not approved for data release to the receiving entity.

7.1.3 The receiving entity shall have a plan in place for securing the data from unintended uses and misuse.

7.2 Data management team will confirm the appropriate variables for export with the GTEx project management team and requesting entity.

7.3 A data export template is developed for review, approval, and versioning of the data export.

7.3.1 Data export templates developed for each receiving entity will be stored on a shared workspace. Data export templates will include general information such as type of export, receiving entity/institution and list of approved recipients, format of export and CDR version.

7.3.2 Additional information such as serology results, toxicology reports or alerts may be inserted, only if previously approved as part of the data export template by the project team.

7.3.3 Modifications to data elements will be controlled by DM, and reviewed by the GTEx project management, approved, and versioned prior to implementation.

7.4 Data is collected in the CDR through direct data entry or any other system or import from any source receiving data for this project.

7.5 DM validates the received data elements through system and manual query resolution.

7.5.1 If there are issues with the received data, DM resolves them working with the BSS, PRC, CBR or other relevant parties.

7.5.2 Only clean data elements, verified as such by DM, are available for export.

7.6 DM validates the data set for export to the receiving entity by confirming the use of the current de-identified data export template that has been approved.

7.7 The project manager or designated staff/entity shall use the approved format for export of the data. This may include emails or electronic data feeds (i.e. XML feeds) directly from the CDR to personnel on the receiving end authorized to receive the data.

8.0 REFERENCE

8.1 HIPAA Privacy Rule: www.hhs.gov/ocr/**hipaa**/finalreg.html