1. 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. 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 (BB) and the Laboratory Data Analysis and Coordinating Center (LDACC).

2.3. Currently all data exports outside of caHUB CDR are deidentified.

2.4. Export to additional entities would require review and approval by the caHUB and GTEx PMT.

3. 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 other 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 caHUB, CDR, data management team, 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. DEFINITIONS

4.1. **Bio4D** - Advanced biospecimen inventory management system provided and administered by the Comprehensive Biospecimen Resource (CBR), and 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.

4.2. **Brain Bank (BB)** – Brain Tissue Core: Receiving facility for the brain, brain stem and hair biospecimens collected by the BSS for the GTEx project.

4.3. **BSS** – Biospecimen Source Site: Institution contracted to identify donors, consent donors, and perform tissue procurement and the associated data extraction for the GTEx project.

4.4. **caHUB -** Cancer Human Biobank operated by Office of Biorepositories and Biospecimen Research- NCI

4.5. **CBR -** Comprehensive Biospecimen Resource: The entity responsible for the receipt, processing, inventory and transfer of GTEx project biospecimens and associated biospecimen data.

4.6. **CDR** - Comprehensive Data Resource: A component of caHUB, the CDR Data Services (DS) is a web-based application and database for all BSS case collection and handling, clinical, pathology and molecular data referenced by Case and Specimen IDs. The CDR-DS is linked to Bio4D and LDACC via web servers to share and transmit such data.

4.7. **Clean Data –** Correction or removal of erroneous (dirty) data caused by contradictions, disparities, keying mistakes, missing information, etc. Data is confirmed clean after review of the data management and /or project management team.

4.8. **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.

4.9. **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 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.

4.10. **Data export** – The formatting and transfer of deidentified data that is used by another application. The export will be based off of the approved Data Export Template and exports a data file in a format that another application understands, enabling the two programs to electronically access the same information.

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

4.12. **Deidentified data** – Data stripped of all the 18 HIPAA identifiers (see HIPAA Privacy Rule).

4.13. **GTEx** – Genotype Tissue Expression project: Initiative designed to determine how genetic variation controls gene expression and its relationship to disease. The National Cancer Institute's (NCI) cancer Human Biobank (caHUB) leads the GTEx biospecimen acquisition sub-initiative comprised of biospecimen source sites, CDR, CBR, PRC and a laboratory data analysis and coordinating center and the Brain Bank.

4.14. **HIPAA** – Health Insurance Portability and Accountability Act

4.15. **LDACC** – The Laboratory, Data Analysis and Coordinating Center: Entity responsible for the overall coordination of GTEx activities, and molecular and statistical analyses.

4.16. **LDS** – Limited Data Set: A limited set of identifiable defined patient information as in the Privacy Regulations issued under HIPAA.

4.17. **MTA/DUA - Material Transfer/Data Use Agreement:** A contractual document used for the acquisition of various biological materials and associated data, specifying the terms and conditions for use and disclosure of the materials and/or data by the providers and recipients of the materials and/or data.

4.18. **PHI** – Protected Health Information

4.19. **PRC** – Pathology Resource Center at caHUB

4.20. **Provider -** Entity providing the data exports to the recipient (for e.g. Data Management team, CDR).

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

4.22. **SOP** – Standard Operating Procedure: Document written to establish the processes and requirements necessary to ensure consistency and complete a given task.

5. ENVIRONMENTAL HEALTH & SAFETY

5.1. Not applicable

6. **MATERIALS/EQUIPMENT**

6.1. Not applicable

7. **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 and it shall include all of the following elements and shall 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 Data Management 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 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 the caHUB SharePoint workspace. Location: ***GTEx Controlled Data Exports*** *(OBBR>caHUB>pilot caHUB>GTEx Workspace>GTEx Controlled Data Exports).*

7.3.2. Data export templates will include general information such as type of export, format of export, receiving entity/institution and list of approved recipients.

7.3.3. Additional information such as serology results or alerts may be inserted, only if previously approved as part of the data export template by the Project Team.

7.3.4. Modifications to data elements will be controlled by the data management team, and reviewed by the GTEx PMT, 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. The data management team validates the received data elements through system and manual query resolution.

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

7.5.2. Only clean data elements, verified as such by the data management team, are available for export.

7.6. The data management team validates the data set for export to the receiving entity by confirming the use of the current deidentified 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 (for e.g. XML feeds) directly from the CDR to personnel on the receiving end authorized to receive the data.

8. REFERENCES

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

9. ATTACHMENTS

9.1. GTEx Deidentified Data Export Process Flow, OP-0008-P1

APPROVALS	
NAME / TITLE	

INITIATION/REVISION HISTORY			
REV #	DESCRIPTION OF CHANGE	AUTHOR	EFFECTIVE DATE