

# The BIG Initiative:

## Frequently Asked Questions for Researchers

### What is the purpose of the BIG Initiative?

- The BIG Initiative (BIG) was created as a means to stimulate biomedical and pediatric healthcare research, support grant applications, and to increase scientific publications by the UTHSC community by facilitating the generation of genomic data. Samples and data that are generated and stored by BIG will be shared with investigators after their requests are approved by the BIG Research Oversight Committee and the UTHSC Institutional Review Board.

### What types of samples are available from BIG?

- BIG isolates and stores genomic DNA from whole blood samples of pediatric patients at Le Bonheur Children's Hospital after informed consent has been obtained. The blood samples we extract DNA from are not dedicated blood draws but rather are scavenged from appropriate blood tubes that are drawn for clinical lab tests ordered as standard of care for each patient. Appropriate blood tubes for DNA extraction include purple top (Na/K EDTA) and light blue top (Na citrate) vacutainer and microtainer tubes. BIG started collecting DNA samples in November 2015.

### How are the DNA samples isolated?

- DNA is extracted from whole blood 7 to 14 days after clinical testing is ordered; this delay is necessary to satisfy Tennessee State Laboratory Board regulations that require blood tubes be held for 7 days prior to disposal to enable any clinical re-testing, if it is required. DNA is isolated in the Le Bonheur Hospital's Molecular Diagnostics Laboratory by an automated QIASymphony S/P system (Qiagen) using magnetic bead-based chemistry. Samples are further processed and archived in the Integrative Genomics Biorepository (IGB), housed in the Le Bonheur Children's Foundation Research Center.

### Do the DNA samples undergo any quality control testing?

- Yes. Isolated genomic DNA samples are tested for DNA concentration by a PicoGreen fluorescent dye-based assay and for median fragment smear size by capillary electrophoresis using a QIAxcel system. DNA concentration is reported in nanogram per microliter (ng/uL) units. Estimated median fragment smear size is reported in base pairs (bp).

### Are the DNA samples standardized to the same concentration?

- No. Because we scavenge what remains in each blood sample, final genomic DNA concentrations vary due to variable blood volumes used in the extractions. Each sample is replicated five (5) times into barcoded storage tubes as is, with each tube receiving 85 uL of DNA in Qiagen elution buffer.

### What if some DNA concentrations are too low for my intended use?

- If some DNA samples have concentrations that are below the recommended levels for your intended use, but you would still prefer to analyze them, we recommend using Agencourt AMPure XP or SPRIselect magnetic bead reagents (Beckman Coulter) to concentrate samples with low DNA concentrations according to the manufacturer's recommended procedures. Alternatively, the QIAamp DNA Micro Kit (Qiagen) can be used. Both methods enable elution of DNA in 10 uL to 20 uL volumes. Please note, however, that Qiagen MinElute Purification kits are not recommended for the concentration of genomic DNA, unless the DNA is sheared, cleaved, or significantly degraded, because they are designed to isolate DNA fragments from 70 bp to 4 kb in size, whereas BIG DNA samples typically have larger median fragment sizes.

## What type of data are available from BIG?

- As the data becomes available, BIG will create an archive of genomic data files that are generated from BIG DNA samples by investigators who analyze BIG samples and by the BIG Initiative itself. Multiple data file types from a variety of genomic analysis technologies may be available, including from next generation sequencing platforms (e.g., FASTQ, BAM, and VCF files) and microarray platforms (e.g., CEL, CHP, GTC, IDAT files).

## Who is eligible to apply for BIG samples and/or data?

- Principal Investigators with a faculty appointment at UTHSC are eligible to apply to BIG.
- Principal Investigators who do not have a primary or joint faculty appointment at UTHSC must establish a collaboration with a faculty member with a primary appointment at UTHSC in order to receive data and/or samples.
- \*Please note\* Principal Investigators with primary appointments at an institution other than UTHSC must receive Institutional Review Board (IRB) approval from their primary institutions along with UTHSC IRB approval in order to receive samples and/or data.

## How do I find out if DNA samples and/or data are available for my study?

- There are four steps to identify appropriate BIG samples and/or data for your proposed study:
  - 1) Principal Investigators must first log into [TriNetX](#) and identify a study cohort by navigating to My Studies > Create New Study.<sup>1</sup> Here, with the use of the Health Data Repository [identifier], TriNetX can show if DNA samples and/or data for patients within that cohort are present in the BIG archive. A cumulative total of DNA samples available for that study cohort will be displayed. A more detailed description of TriNetX(rEDW) can be found at CBMI Lab > [TriNetX Live](#). Please call CBMI at 901-287-5834 or email [Ichintha@uthsc.edu](mailto:Ichintha@uthsc.edu) with any questions about TriNetX access, training, and use.
  - 2) If this number of samples are of interest to the Principal Investigator, the study can be shared with Lokesh Chinthala in TriNetX in order to obtain a list of DNA samples and/or data files available for that study cohort.
  - 3) BIG will contact the Principal Investigator by email to ensure the search parameters used in TriNetX are precise and accurate for the intended study, and will additionally use the Research Enterprise Data Warehouse (rEDW) to search Methodist Le Bonheur's Cerner EMR to ensure the cohort is fully up-to-date.
  - 4) BIG will return a list of corresponding clinically valid samples with quality control metrics and/or data files to the Principal Investigator.<sup>2</sup> This BIG List can be used as is, or edited by the Principal Investigator to remove undesired samples, to create a final list prior to formal application to request the samples and/or data.

<sup>1</sup>*TriNetX is maintained by the UTHSC Center for Biomedical Informatics (CBMI) and contains a variety of de-identified clinical and demographic information for all patients of Methodist Le Bonheur Health System.*

<sup>2</sup>*Because BIG and TriNetX are both de-identified archives, BIG works with CBMI-certified honest brokers to match TriNetX patient cohorts with BIG samples and/or data. Identifiers in BIG and TriNetX lists and files are randomly generated and do not contain any protected health information (PHI).*

## How do I request DNA samples and/or data from BIG?

- After receiving the BIG List of samples and/or data that are relevant to your proposed study, complete the following three steps to make a formal request for samples and/or data:
  - 1) Fill out a [BIG Materials Distribution Request Application](#).
  - 2) Provide and/or upload all information and documents requested in the application. A description of all required information is provided below.
  - 3) Submit the completed application. You can save incomplete applications until ready for submission to BIG.

## What information is requested in the BIG Materials Distribution Request Application?

- Information on the Principal Investigator and any project collaborators:
  - PI name
  - PI department/affiliation
  - PI campus/institution location
  - PI contact info
  - Names and affiliations of project collaborators.
- Descriptive sections about the project and its scientific merits:
  - Project Narrative suitable for the general public (1-3 sentences) – BIG Newsletters to its participant community may include descriptions of the research BIG samples are used for.
  - Project Summary (300 words or less).
  - Impact Statement on Healthcare (150 words or less).
  - Scientific Significance of Project (150 words or less) - should include significance/importance of BIG samples/data to project.
  - Research Approach (150 words or less) - should include description/rationale of data generation platform (e.g., whole exome sequencing, GWAS, etc.).
  - Sample number and power estimate based on sample size; or justification for a low powered cohort.<sup>3</sup>
  - Data analysis plan (200 words or less) - should include name and bona fides (e.g., website URL, contact information, etc.) of the data-generating service provider.
  - Plan for sample and data security if samples/data are to be sent outside UTHSC (excluding summary data reported in scientific publications/presentations).<sup>3</sup>

<sup>3</sup>UTHSC resources available for questions on power calculations and data security include:

[Biostatistics, Epidemiology and Research Design \(BERD\) Clinic and Online Resources](#)  
[Center for Biomedical Informatics](#)  
[Department of Preventive Medicine](#)

- Funding source(s) for data generation – assured data generation is an important BIG goal:
  - Industry: Name of sponsor
  - Previously Funded Grant: Name of funding agency
  - Proposed Grant: Name of funding sponsor and submission deadline/date
  - Institutional/Departmental funds: Name of Department and Department Chair
- Status of peer-review of your project:
  - NIH: review complete/under review
  - Departmental: review complete/under review
  - Other peer review: review complete/under review
  - Not yet peer-reviewed: Describe plan, if any, for obtaining peer-review (50 words or less)

- Estimated date (year/month) when you need the samples/data.
- Description of ethical concerns or issues that could be raised by the project (150 words or less).

### **What electronic documents should accompany the BIG Materials Distribution Request Application?**

- Quotations from the data-generating service provider (if using one). This includes the Molecular Resource Center (MRC).
- If institutional/departmental funding is being used, a letter from the Department Chair that corroborates availability of the funds.
- Principal Investigator's NIH Biosketch with Funding Support page.
- The BIG List: excel format file with sample identifier list for the requested samples and/or data for the proposed study cohort.

### **How will my BIG Materials Distribution Request Application be reviewed?**

- Submitted applications are electronically distributed on a monthly basis to five (5) members of the BIG Research Oversight Committee (ROC) for initial review and disposition.
- Email notification of unanimously approved requests will be passed on to the IGB and the Principal Investigator.
- Applications given any negative decisions (denials) from ROC reviewing members will go to a monthly ROC meeting for group discussion and a vote. Simple majority rules.
- An invitation to attend the ROC meeting will be issued to the Principal Investigator of any application under group review. The PI will be given the opportunity to give a brief 5 to 10 minute presentation to the ROC to support the request as well as to address questions from the ROC.
- Official ROC approval or denial letters will be sent to the Principal Investigator within one month of the ROC meeting. ROC Approval Letters should be included with IRB applications and can be used as supporting documents for grant applications.

### **What Decision Criteria are used by the ROC?**

- UTHSC affiliation:
  - First priority will be given to PIs whose primary faculty appointment is at UTHSC.
  - Second priority will be given to faculty with a primary appointment at St. Jude Children's Research Hospital and a joint appointment at UTHSC.
  - Third priority will be given to UTHSC faculty with a primary appointment at another UT campus and a joint appointment at UTHSC.
- Complete Application: All requested information is provided.
- Impact on Healthcare: Rate the impact on healthcare of the proposed project using the NIH 9 point scoring system (1 = highest impact; 9 = lowest impact)
- Scientific Significance: Rate the significance of the proposed project using the NIH 9 point scoring system (1 = highest significance; 9 = lowest significance)

- Ethical considerations: If the PI describes potential ethical issues, or a ROC reviewer determines that the proposed project poses potential ethical issues, the application will be routed to the BIG Ethics Committee for evaluation of those issues and to render a determination as to whether those ethical issues preclude approval.
- Sample Number and Power Estimate: Sample number and power calculations to estimate the appropriate/minimal sample sizes required for a statistically significant result will be afforded wide latitude by the ROC due to the fact that all sequence and genotyping data generated from BIG samples will be returned to the BIG data archive, as mandated by the BIG Research Materials Use Agreement, for future use by UTHSC investigators. In other words, the BIG Initiative would like to collect as much genomic data as possible. However, the academic products derived from BIG samples and/or data (e.g., publications, funded grants, etc.) will reflect the success of the BIG Initiative. Moreover, given the limited opportunities for sample accrual, BIG must view its DNA samples as a non-renewable resource. Therefore,
  - Highest priority will be given to projects for which power estimates suggest a statistically significant result can be obtained.
  - In the absence of that determination, priority will be based on the number of requested samples, with greater priority given to greater sample numbers.
  - Projects that request/require more than one vial of DNA per sample or the last vial of a sample will be judged with greater scrutiny due to the potential depletion of samples in the BIG archives.
- Genomic data generation methodologies proposed for samples:
  - The generation of data that could be used by any/all UTHSC investigators is part of the BIG Initiative's mandate. Therefore, highest priority will be given to high-throughput data generation methods that generate the greatest amount of data per sample; e.g., (in descending order):
    - i) Whole Genome Sequencing (30X average depth or greater)
    - ii) Whole Exome Sequencing (50X average depth or greater)
    - iii) Long Read Genomic Sequencing (10X average depth or greater)
    - iv) Genome-wide SNP genotyping (900K SNPs or greater)
    - v) Genome-wide CNV genotyping (850K markers or greater)

This list will be amended as better, more efficient methodologies become available.
  - It is also recognized that more targeted approaches can have significant value. Therefore, requests involving targeted gene panels for deep resequencing or low to mid-range throughput SNP analysis will be given mid-priority following ROC consideration on a case-by-case basis.
  - No/low priority will be given to other small data generation projects (e.g., single or several SNPs, initial sequencing of single genes or small targeted NGS panels, etc.) based on their low impact relative to BIG costs.
- Data-generation Service Provider: The name and bona fides (e.g., website URL, contact information, etc.) of the facility to be used for data generation and a quotation for analysis of the proposed number of samples must be provided by the requesting Principal Investigator. Sample distribution will not be approved without such proof.
- Funding Source: Proof of sufficient funds dedicated to completion of the data generation aspect of their project must be provided by the requesting PI. Sample distribution will not be approved without such proof.

## **After the ROC approves my request, what other steps are needed to receive samples and/or data?**

- Submit ROC approval letter with your study protocol application to the IRB.
- Receive IRB approval for your study protocol.
- Provide a copy of the IRB approval letter with protocol number and approval date to BIG ([biglist@uthsc.edu](mailto:biglist@uthsc.edu)). BIG will email the Research Materials Use Agreement to you.
- Sign and return the BIG Research Materials Use Agreement to BIG ([biglist@uthsc.edu](mailto:biglist@uthsc.edu)).
- BIG will contact you (the Principal Investigator) when the samples and/or data are ready for distribution.

## **In what time frame must sample data be generated and returned to the IGB?**

- As stipulated in the BIG Research Materials Use Agreement:
  - Data should be generated for all distributed samples within one (1) year of receipt of BIG samples by the Principal Investigator.
  - Copies of electronic data files generated from BIG samples should be received by the IGB within one (1) year of receipt of BIG samples by the Principal Investigator. These files should enable the identification, extraction, and re-analysis of data for each and every distributed and analyzed BIG sample.
  - You can contact BIG via email ([biglist@uthsc.edu](mailto:biglist@uthsc.edu)) to arrange data file transfers.

## **What if I do not get the BIG samples analyzed as proposed in my application?**

- As stipulated in the BIG Research Materials Use Agreement:
  - Any BIG sample that is not analyzed within one (1) year of receipt by the Principal Investigator must be returned to the IGB.
  - After one year, the IGB will contact you regarding the disposition of the BIG samples and data files generated from their analysis. Arrangements for the return of the samples to the IGB can be made at that point in time.
  - You can contact BIG via email ([biglist@uthsc.edu](mailto:biglist@uthsc.edu)) to arrange sample returns.

## **What data file types should I return to the IGB?**

- Data file types for each distributed sample analyzed on Illumina or Ion next-generation sequencing platforms include:
  - \*.FASTQ
  - \*.BAM
  - \*.BAM.BAI
  - \*.VCF or \*.GVCF
- Data file types for each distributed sample analyzed on Illumina microarray platforms include:
  - \*.GTC
  - \*.IDAT
  - \*.TXT (for Beta values, p-values, signal\_A values, signal\_B values, and the sample table)
- Data file types for each distributed sample analyzed on Affymetrix microarray platforms include:
  - \*.ARR
  - \*.CEL
  - \*.CHP