TREC Biobanking and Clinicogenomics Protocol

Specimen and Data Management

Consent will be obtained on paper by a clinical research staff member, nurse, pharmacist, or physician. A licensed nurse or trained phlebotomist will draw the blood samples at enrollment. Four 8.5 mL (PAX RNA, PBMC, Plasma) and one 2.5 mL (PAX DNA) blood samples will be collected (Table 2; approximately 35 mL). No visits beyond the initial visit in which the patient consents and provides biospecimens are required. However, a 10 μL (a small drop comparable to a blood glucose test) finger prick sample using the self-administered Mitra device[51] (see Stanford human subjects materials) will be requested every 6 months for the biobank during the 5 year participation period (up to 10 samples total). The Registry and Protocol will be posted on ClinicalTrials.gov.

Biospecimen protocol
Multiple biospecimens (Table 2 below) will be collected and banked for future research (Figure 2, not highlighted). At the point of collection, samples will be assigned a unique code for each participant that is linked to data abstracted from the EMR (data are coded). All data (each element), will have an attached code for that patient in meta-data enabled by the Burst IQ digital ledger linking the case level data (but lacking 45 CFR §164.514 direct identifiers) with case level data lacking the enumerated as specified in the Clinical Curation section below, to the sample. Samples and data will be linked using a BarTender barcoding system that connects the clinical annotation data, metadata on the BurstIQ digital ledger, and the printed cryolabel on each specimen vessel. The consent form (whether electronic or digital) is attached to each code, so each case level data set is coded with metadata that documents patient consent with ground truth (a copy of the signed consent form) and associated permissions. A valid consent is verifiable and visible to TAMU staff with a regulatory role for every case level data set in the BurstIQ platform.

Clinical Samples: No clinical testing is required by this protocol, but any results from such testing for medical reasons under applicable standards of care will be extracted from the medical record and included in the research data set.

Biobank (Research) Samples: For general purposes, clinical staff will draw blood into each vacutainer tube in Table 2 under gloved and alcohol clean conditions to minimize the risk of infection. Personal protective equipment will be used in accordance with established clinic safety protocols. Long term storage of specimens will be at the Texas A&M Institute of Bioscience and Technology (IBT) as described in section 9. Following immediate collection of samples, specimens will be temporarily stored at the Family Medicine Clinic in a dedicated cold storage refrigerator under the supervision of trained staff at 4°C until transport and processing for temporary storage at the Reynolds building on the TAMU College Station campus by staff with appropriate training. Transport from the clinic to Reynolds will comply with the guidelines established in the TAMU Biosafety Manual for Risk Group 2. Samples will be aliquoted into smaller quantities in a laboratory with biocontainment level 2 and then placed in liquid nitrogen tanks before scheduled shipping to the IBT for long-term storage in liquid nitrogen, as outlined in the addendum in Table 2. Shipment to Houston will be performed weekly by a FedEx on dry ice per their guidelines:

https://www.fedex.com/content/dam/fedex/ca-canada/MVP/images/2020/Q2/FX_HowToPack_Clinical_CA_EN_1248052062.pdf .

Research specimens will be kept at Texas A&M's secure biorepository located at:

Institute of Biosciences and Technology
2121 W. Holcombe Blvd.
Houston, Texas 77030.

The IBT biobank is a dedicated room on the 1st floor of the Alkek Building that has keycard access and backup generator power. Only study staff and their designees have access to the specimens in the IBT Biobank. Monitoring, refrigeration, or freezing of samples will be available as needed.

All samples will be aliquoted into 500mL cryotubes to minimize freeze-thaw cycles. Each sample will be labeled with a patient ID number and barcode printed cryolabel to enable matching with relevant case level data from biobank activities approved by the Institutional Review Board (IRB). However, the codes required to reidentify the samples will only be available to study staff with access to patient data (Qiang He, DeLona Bacote), or through a tokenization service available from BurstIQ. The syntax for the human readable coding will specify the location, collection date, coded patient ID#, and type of collection medium (i.e., plasma with EDTA vacutainers). The BarTender barcode system has been implemented at the IBT Biobank and barcodes for specimens will be integrated with case level research data in the BurstIQ meta data ledger that omits 45 CFR §164.514 direct identifiers.

Whole Genome Sequencing (WGS) will be conducted on Illumina HiSeq at 30X coverage by Admera, the company that has done single cell sequencing for the TREC Single Cell Core. Data from WGS will be uploaded to TAMUs AWS cloud enclave. **WGS data will be for research use only** and will not be reported to prescribing practitioners or participants. These data will be used in accordance with the section on research data government in the biobanking addendum. Incidental findings will not be shared with participants under this protocol. WGS data will only be accessible to the investigators listed herein or third-party research collaborators in accordance with IRB oversight of future proposed studies. Consent forms will be signed with a wet ink signature on a printed consent form or as a secure digital signature on the PDF file using the Adobe Sign function and attached to the PROVENANCE and WGS metadata using a script developed in the BurstIQ platform.

IRB NUMBER: STUDY2024-0008
IRB APPROVAI DATE: 11/6/2024

Table 2. Biospecimen Collection for TREC Registry

Size	Collection Media	Sample Type	Clinical Use Regulatory Context	Use Case	Bioanalyte	Bioanalysis Method	Methodology Protocol Reference	Storage
120mL	Sterile Sample Cups	Urine	LDT	Clinical/ CLIA lab	Drug levels	MassSpec		Direct to Lab
N/A	Cotton Tipped Applicator	Buccal Swab	LDT	Clinical/ CLIA lab	PGx	ProbeArray		Direct to Lab
10 µL	Neoteryx Mitra	dried bloodspot	LDT/RUO	Biobank	TBD	TBD	Shen, X., Kellogg, R., Panyard, D.J. *et al. Nat. Biomed. Eng* (2023).	liquid N_2
8.5 mL	PAX RNA	Blood RNA	RUO	Biobank	TBD		Donohue DE, Gautam A, Miller SA, Srinivasan S, Abu-Amara D, Campbell R, Marmar CR, Hammamieh R, Jett M. PLoS One. 2019 Oct 10;14(10):e0223065.	liquid N_2
2.5 mL	PAX DNA	DNA	RUO	Biobank	TBD		Voss T, Ullius A, Schönborn M, Oelmüller U (2021) PLOS ONE 16(7): e0253401.	liquid N_2
8.5 mL	BD P800 Blood Collection System	Stable Plasma Proteins	RUO	Biobank	TBD	Proteomics	Debunne N, De Spiegeleer A, Depuydt D, Janssens Y, Descamps A, Wynendaele E, De Spiegeleer B. ACS Omega. 2020 Jun 22;5(26):16120-16127.	liquid N_2
N/A	Cotton Tipped Applicator	Buccal Swab	RUO	Biobank	WGS/WES	NGS		liquid N_2
N/A	Thermo Scientific™ SpeciMAX™ Saliva Collection Kit	Saliva [Cost dependent]	RUO	Biobank	TBD	TBD		liquid N_2
8.5 mL	Vacutainer CPT Red/Green	PBMC	RUO	Biobank	TBD	RNA seq	Betsou, F., Gaignaux, A., Ammerlaan, W. *et al. Curr Pathobiol Rep* **7**, 17–27 (2019).	liquid N_2
8.5 mL	Vacutainer Purple top	Plasma	RUO	Biobank	TBD	TBD		liquid N_2
8.5 mL	Vacutainer SST Red/Gray	Serum	RUO	Biobank	TBD	TBD	VJ Bush, MR Janu, F Bathur, A Wells, A Dasgupta. Clinica Chimica Acta. Volume 306, Issues 1–2. Pages 139-143, (2001).	liquid N_2
8.5 mL	Vacutainer Yellow Top	Whole Blood	RUO	Biobank	TBD	TBD		liquid N_2

IRB NUMBER: STUDY2024-0008
IRB APPROVAI DATE: 11/6/2024

Follow-up Remote Sample Collection

Participants will be offered the opportunity to submit follow-up samples near the 6-month and 1-year anniversary of their TREC enrollment visit for 5-years (see Figure 1), under the same scope of consent bioanalysis as the original IRB approved informed consent form referenced herein. The Neoteryx Mitra micro-sampling (10 μL) technology will be utilized for home-based participant self-collection of dried bloodspots.[48] Dried bloodspots were used for MTM in transplant patients during COVID-19 lockdowns to prevent exposure of immunocompromised patients in healthcare facilities. The Mitra system has been used for assessment of blood levels of therapeutic antibodies (30 days at room temperature),[49] transplant drugs (96 days up to room temperature),[50] and multi-omic analysis[51] The Mitra device is FDA class 1, CE marked in Europe, and can be used in compliance with most CLIA lab developed tests.

Clinical Data

An authorized clinical user (e.g., a member of the covered entity workforce) can log in with a username and password to access a participant's data when warranted. The clinical annotation and phenotypic data (information obtained in examinations and physician consultations, excluding genetic data) will be collected on-site by the clinical team and entered in eCW through a cloud portal to a data center. Research data will be abstracted from eCW and ActX and stored and managed on a platform using BurstIQ's data network. All software and IT services vendors for the Clinicogenomics Core have completed a Security Assessment by TAMU IT (eCW, ActX, BurstIQ, AWS), and TAMU IT can provide details. A digital copy of the subjects' consent form, and metadata if electronic (i.e., AdobeSign), is attached to the PROVENANCE blockchain for that participant to ensure the provenance of consent for each participant-level dataset for authorized personas (only study staff).

The frequency of data refresh is estimated to be once per quarter across the cohort.

Specimen and Bioanalysis Data Sharing

The data to be collected during the study are enumerated in section 9 and Table 1.

Clinical Curation

During sample collection, each sample will receive a unique code that is linked to patient data from the EMR. These data are coded and contain no direct identifiers as specified in 45 CFR §164.514. The case level data, along with clinical annotations, is connected to the sample using a BarTender barcoding system. Additionally, the BurstIQ digital ledger stores metadata linking the case level data to the sample. Each sample vessel is labeled with a printed cryolabel that also contains metadata. The patient consent form, whether electronic or digital, is linked to each code, documenting patient consent and associated permissions. TAMU staff with regulatory roles can verify and access valid consent for every case level data set in the BurstIQ platform.

Specimens will be affixed with a cryolabel that encodes the following information:
Example: 1001-072123-XD-BCSHH = GUID - collection date- sample type- location

Specimen Types:
Sr-Serum | Pl-Plasma | Bl- Blood | U- Urine | Bs-Buccal Swab | S-Saliva | XD- PAX DNA XR-PAX RNA

Collection Location:

BCSHH- Bryan- College Station Health Hub | Navasota Clinic

Certain data generated during clinical care (Real Word Data; RWD will be captured in the respective Clinical EHRs (eCW for Family Medicine and the ActX for patients that happen to also visit the IPGx) and transferred to the Research Data Warehouse. These data sets will be extracted, abstracted, annotated, and replicated in the PROVENANCE Research Data Warehouse and digitally linked to each specimen using a unique identification number and a bar code such that deidentified case level annotation is available for future research subsequent to bioanalytic examination of samples.

- Claims and billing data
- Chief complaints
- Progress and procedural notes
- Demographics (age, race, zip code, sex)
- Medication history
- Diagnoses
- Hospitalizations
- Physician notes
- Pharmacy
- Vitals
- Continuity of care and Clinical Data Architecture documents, including but not limited to
    - OVS – Office visit summary document
    - VDT – View download transmit summary
    - TOC—Transfer of Care ambulatory summary
- Self-reported or documented ADRs. Documented ADRs will be classified in accordance with NCI USNCI Cancer Therapy Evaluation Program: Common Terminology Criteria for Adverse Events (CTCAE) v3.0. Available online: https://ctep.cancer.gov/protocoldevelopment/electronic_applications/docs/CTCAE_v5_Quick_Reference_8.5x11.pdf (accessed on 4 February 2020).

The following data elements are expected to be collected and entered in the PROVENANCE Registry Data Warehouse:

Labs

- .vcf file
- Clinical chemistry
- Drug concentrations
- ***Note: Labs are not required by the protocol, but data will be shared pursuant to the patient consent if that data becomes available***

The PROVENANCE Registry Clinical Research Data Warehouse will collect clinical EMR data from the Family Care Clinic and will also generate its own data using specific disease and quality of life scales outlined in the protocol. The PROVENACE Data Warehouse will receive research data, but not HIPAA data, as the patient signs a HIPAA authorization form with consent from the Family Medicine Clinic. The Health Level Seven International (HL7) Clinical Document Architecture (CDA) documents will be imported into the research data warehouse.

HL7 is a data standard used to transfer data between electronic medical records to ensure data continuity during referral care. It can also be utilized to overcome interoperability barriers in extracting EMR data for research. Clinical data from participating patients will be periodically curated and abstracted from the EMR and Clinical Decision Support (CDS) systems (if used for a specific patient) into the PROVENANCE Research Data Warehouse, under the patient's authorization, HIPAA waiver, and informed consent as outlined in the protocol, subsequently being classified as Research Data.

Research Data Governance

The TREC REGISTRY Data will include personally identifiable information (PII) and will be stored in compliance with applicable state and federal privacy regulations and principles, including HIPAA authorization. It will only be accessible to the investigators listed in the document or third-party research collaborators in accordance with IRB oversight of future proposed studies. The BurstIQ platform enables a ledger of future data uses to be readily available for compliance audits.

Research data will primarily be stored on the BurstIQ platform or in Aggie Star AWS enclaves. Security, backup, encryption, and authentication are managed in TAMU's Secure Data Zone on the platform. A security assessment has been conducted by TAMU Information Technology. The BurstIQ platform allows granular role assignments, ensuring that each user will have access only to the content, functions, or data required to perform their role. Researchers will only be able to access and use data that they have been authorized to see and use. Likewise, staff and employees will only be able to access and edit functions and content relevant to their roles. Data cannot be downloaded from the platform without proper authorization.

Even though data from EMRs become research data with participant HIPAA waiver, the data sets will be de-identified at the case level with HIPAA identifiers removed when shared from the Research Data Warehouse on the Aggie STAR platform on AWS and to an existing external data management platform (BurstIQ) that uses AWS servers in Virginia and Oregon. Case-level (de-identified) data annotated to genotype data will be considered part of the TREC Clinicogenomic Registry. The TREC Clinicogenomic Registry dataset is a valuable collaborative currency of the TREC program and is intended to be shared with external third parties in the medical research community (including industry) in a de-identified format, in accordance with third-party obligations, patient consent, and applicable privacy and data security regulations. TREC Research Data will be de-identified at the case level but linked to an identifier, making it possible to go back into the medical record or re-consent the patient for future research.

Under this protocol, the TAMU Health Compliance Officer will have review and release oversight for data and specimen requests and will refer to other TAMU units as warranted. Non-investigator collaborators will be required to have IRB review or a waiver of human subject's research, which must be reviewed by the Texas A&M Human Subjects Office (or the institutional HSPO/IRB for an external academic collaborator, and a central IRB for an industry collaborator) prior to data or specimen access. Patient consent and IRB approval will be digitally enforced on the Provenance Platform, ensuring that collaborator access to data cannot be granted until all requirements under this protocol are met, including administrative review of outbound data batches to confirm de-identification.

The purpose of this research exercise is to conduct case-level analysis of healthcare utilization corresponding to disease burden data such as ICD-10 codes, medication utilization, and

specialty care outside the Texas A&M Health System. Genomic research funded by the NIH requires upload to relevant NIH genomic and clinicogenomic databases. When required by the sponsor, the minimum required de-identified data will be uploaded if patient consents allow.

Limited Reidentification

The TREC@TAMU Cancer Prevention Clinicogenomic Registry Data is intended to provide a holistic picture of a patient's health journey, including their interactions with societal, environmental, occupational, and financial factors that influence health outcomes. This data will be deidentified at the case level but linked to an identifier, allowing access to the patient's medical record or the possibility of reconsenting the patient for future research. Through informed consent, researchers will be granted permission, via a limited HIPAA waiver, to match the patient's case-level medical record data with deidentified claims data from available clearinghouses, such as the All-Payer Claims Database (APCD) and the Greater Houston Health Connect, using probabilistic or deterministic data matching methods.[47] A designated TAMU official or honest broker will manage the identifier. The informed consent will also allow limited reidentification of samples by the investigators for the purpose of obtaining claims data from commercial claims databases (like the All-Payer Claims Database or TX Medicaid databases) to engage in probabilistic or deterministic matching (34, 35) for assessing healthcare utilization and health outcomes that can be gleaned from health data. Research activities involving case level data across the health ecosystem will require external data (e.g., medical claims) and expertise. Specific project protocols will need to undergo IRB review.(35)

Specimen sharing

Specimens are intended to be shared in collaborative research projects with high scientific merit and in accordance with the scope of consent provided by participants. Research use of specimens by investigators on this protocol must be authorized by the PI (Dr. Ramos and the Texas A&M IRB). Research use of coded specimens by investigators other than listed investigators on this protocol must be authorized by the PI (Dr. Ramos) and the Texas A&M IRB, as warranted under law and university policy. The BurstIQ platform enables a digital ledger coupling coded samples with case level data so reidentification does not need to occur outside the computational sandbox of the BurstIQ platform. Storage and use of TREC data that includes PII will be conducted in strict accordance with the informed consent instruments herein, the approval of the IRB, applicable statutes governing the privacy of health data, and associated third party contractual obligations. The HIPAA authorization for participant data results in no PHI being in the PROVENANCE data warehouse.

Sharing of data or specimens, by TAMU, with a third party may require the following steps:

- Written approval from TAMU officials (to trigger an MTA and/or DUA), PI Authorization, and if necessary ICF verification (which BurstIQ platform provides digitally for every bit of data)
- Execution of a specimen transfer and data use agreement that ensures HIPAA/GDPR compliance and compliance with the informed consent and this Protocol.
- Specifies disposition of residual specimen, genetic material, and data upon completion of usage or expiration of the underlying agreements and permissions (e.g., destruction, return).
- Strictly prohibits secondary or tertiary transfer or sharing except for bioanalytic work by contractors under strict control of research agreements and this protocol.

Delineation of research and clinical activities

Data collection is for research purposes only. No medical interventions or clinical testing are part of the protocol. Any bioanalytic data generated from TREC biobank samples (such as DNA sequencing, proteomics, transcriptomics, and metabolomics) will only be used for research and will not be included in the medical records of participants under any circumstances. This information will not be available for use by healthcare providers or patients. The study is purely observational in format.