OOR Data-Access Proposal — dbGaP Controlled Data

**Project: TCGA-GBM (phs000178) — WGS & WXS**

Institution: The Ohio State University (OSU)

Department/Center: Radiation Oncology, OSUCCC – James

Today’s date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# 1) Project Overview

**Title: APOE genotype distribution and clinical association in TCGA-GBM using whole-genome and whole-exome sequencing**

Datasets: TCGA-GBM (phs000178), controlled access

Modalities requested: Whole-Genome (WGS); Whole-Exome (WXS)

Access route: dbGaP authorization → GDC Data Portal (NIH login + token)

Purpose (high level):

* Estimate the distribution of APOE genotypes (defined by canonical APOE SNP haplotypes) in GBM patients.
* Evaluate associations between APOE genotype and clinical outcomes (e.g., overall survival), reporting aggregate, non-identifiable statistics only.

**Why germline information is required:**

* APOE genotype is defined by germline SNPs (e.g., rs429358 and rs7412). Accurate determination requires normal/germline sequencing where available.
* Tumor genomes may exhibit copy-number changes, loss of heterozygosity (LOH), or somatic artifacts that can bias genotype inference; matched germline avoids these confounders.
* Linking APOE germline genotype to survival and other clinical endpoints requires reliable per-patient germline calls; results will be reported only at the cohort/aggregate level.

# 2) People & Roles

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Name | Role | eRA Commons ID | Email | Phone | Access Group(s) | Start–End |
| Deliang Guo, PhD | PI / Approved User | \_\_\_\_\_\_\_\_\_\_ | \_\_\_\_\_\_\_\_\_\_ | \_\_\_\_\_\_\_\_\_\_ | PI‑Secure | \_\_\_\_\_\_\_\_\_\_ |
| Zijie Feng | Project Lead / Authorized Downloader | \_\_\_\_\_\_\_\_\_\_ | \_\_\_\_\_\_\_\_\_\_ | \_\_\_\_\_\_\_\_\_\_ | dbGaP‑Downloaders | \_\_\_\_\_\_\_\_\_\_ |
| \_\_\_\_\_\_\_\_\_\_ | Analyst / Authorized User | \_\_\_\_\_\_\_\_\_\_ | \_\_\_\_\_\_\_\_\_\_ | \_\_\_\_\_\_\_\_\_\_ | Analysts‑Secure | \_\_\_\_\_\_\_\_\_\_ |
| Michael | IT/Security Contact (Sr. Director of IT) | N/A | \_\_\_\_\_\_\_\_\_\_ | \_\_\_\_\_\_\_\_\_\_ | Security‑Admins | \_\_\_\_\_\_\_\_\_\_ |
| \_\_\_\_\_\_\_\_\_\_ | Institutional Signing Official (SO) | N/A | \_\_\_\_\_\_\_\_\_\_ | \_\_\_\_\_\_\_\_\_\_ | SO | \_\_\_\_\_\_\_\_\_\_ |

# 3) Research Use Statement (final)

We request controlled-access TCGA-GBM (phs000178) WGS/WXS files to (1) derive APOE germline genotypes (based on canonical APOE SNP haplotypes) and quantify their distribution in GBM patients, and (2) assess associations with clinical outcomes (e.g., overall survival), reporting aggregate, non-identifiable results only. Work will be conducted solely for biomedical research consistent with the dataset’s Data Use Limitations and the current Data Use Certification. Controlled data will be stored and analyzed only within an OSU NIST SP 800-171–aligned secure environment with role-based access, MFA, encryption in transit/at rest, and audit logging; data will be retained only for the approved period and destroyed at close-out with SO confirmation.

# 4) Specific Aims

1. Genotype & distribution: Call APOE germline genotypes from WGS/WXS (prefer normal DNA where available) and tabulate genotype/allele frequencies across the cohort.
2. Clinical linkage: Link APOE genotype to clinical metadata and perform survival analyses (Kaplan–Meier, Cox proportional hazards) with available covariates (e.g., age, sex, treatment variables as available).
3. Reporting: Release only aggregate statistics (e.g., hazard ratios, survival curves, frequency tables); no individual-level data.

# 5) Methods (summary)

* Inputs: Controlled WGS/WXS (normal and/or tumor), plus TCGA clinical files (survival/time-to-event, vital status, basic demographics).
* Variant calling for APOE: Extract germline genotypes for the canonical SNPs used to define APOE haplotypes/isoforms; confirm call quality/coverage.
* Quality control: Sample identity checks, depth/quality thresholds, concordance across WGS/WXS if both exist.
* Statistics: Genotype counts and allele frequencies; survival analyses (KM/Cox), sensitivity analyses, covariate adjustments as available.
* Outputs: Aggregate tables/figures only; no individual-level genotypes or raw reads leave the secure environment.

# 6) Human Subjects / IRB

Does this access require an OSU IRB approval/acknowledgment?

* [ ] Yes — attach IRB letter (Protocol #: \_\_\_\_\_\_\_\_\_\_)
* [ ] No / Not required (controlled, de-identified external data; adhering to NIH GDS & DUC)
* [ ] Unsure — please advise

OSU HRPP contact (if needed): \_\_\_\_\_\_\_\_\_\_

# 7) Security & Compliance (one-page summary)

* Secure environment name/location: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
* Standards: NIST SP 800-171 alignment for storage/compute of controlled human genomic data
* Identity & access: Institutional SSO + MFA; least-privilege groups for authorized users only
* Encryption: At rest (AES-256 or equivalent); in transit (TLS 1.2+)
* Network/segmentation: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
* Logging & monitoring: Authentication, access, and file operations auditing; retention: \_\_\_\_\_\_\_\_\_\_
* Patching/vulnerability management: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
* Data sharing: No sharing of controlled files outside approved environment; no re-identification
* Backups: Encrypted; scope/retention: \_\_\_\_\_\_\_\_\_\_
* Data destruction: Verified secure deletion at close-out; SO confirmation recorded

# 8) Operational Plan (Data Flow & Procedures)

Acquisition:

1. dbGaP DAR approved → NIH login to GDC
2. Filter Project = TCGA-GBM; Experimental Strategy = WGS, WXS → export manifest
3. Download using gdc-client with user’s GDC API token into encrypted landing storage

Processing & Analysis:

* Ingest to controlled compute workspace (same secure environment)
* Derive APOE germline genotypes; perform QC; link to TCGA clinical (survival)
* Run survival analyses (KM, Cox); generate aggregate tables/figures
* Intermediate & final files remain within enclave; no public/cloud egress of individual-level data

Outputs: Only aggregate, non-identifiable results (frequency tables, HRs, KM plots) leave the enclave.

Lifecycle: Access window = 1 year from approval; Renewal before expiry or close-out with verified deletion and SO confirmation.

Logical data-flow diagram:

GDC (NIH login + token)  
 └─> Encrypted Landing Storage (OSU Secure Env)  
 └─> Controlled Compute Workspace (APOE genotyping + survival)  
 └─> Aggregate Results (tables/figures only)  
 └─> Publication / Presentation

# 9) Retention, Renewal & Close-Out

* Planned retention during approval window: \_\_\_\_\_\_\_\_\_\_
* Annual report/renewal needed by: \_\_\_\_\_\_\_\_\_\_
* Close-out deletion method/tooling: \_\_\_\_\_\_\_\_\_\_
* SO confirmation method (ticket/email/form): \_\_\_\_\_\_\_\_\_\_

# 10) Training & Policies

* Team acknowledges/has completed required OSU training for restricted data handling: [ ] Yes [ ] No [ ] In progress
* Applicable internal policies followed (list): \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ (e.g., Research Data Policy, Institutional Data Policy, Secure Computing Guidelines)

# 11) Attachments (check & include)

* [x] Research Use Statement (Section 3)
* [x] Specific Aims & Methods (Sections 4–5)
* [x] Security summary (Section 7)
* [x] Operational plan & data-flow (Section 8)
* [ ] IRB letter (if required)
* [ ] Other OOR/OSP forms: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# 12) Routing & Timeline

* OOR/OSP submission date: \_\_\_\_\_\_\_\_\_\_
* IT attestation call (Michael) date/time: \_\_\_\_\_\_\_\_\_\_
* dbGaP DAR submission date: \_\_\_\_\_\_\_\_\_\_
* DAC decision date (est.): \_\_\_\_\_\_\_\_\_\_
* Project start (post-approval): \_\_\_\_\_\_\_\_\_\_

# 13) Points of Contact

* PI: Deliang Guo, PhD — Email: \_\_\_\_\_\_\_\_\_\_ — Phone: \_\_\_\_\_\_\_\_\_\_
* Project Lead: Zijie Feng — Email: \_\_\_\_\_\_\_\_\_\_ — Phone: \_\_\_\_\_\_\_\_\_\_
* IT/Security: Michael (Sr. Director of IT) — Email: \_\_\_\_\_\_\_\_\_\_ — Phone: \_\_\_\_\_\_\_\_\_\_
* Signing Official (SO): \_\_\_\_\_\_\_\_\_\_ — Email: \_\_\_\_\_\_\_\_\_\_ — Phone: \_\_\_\_\_\_\_\_\_\_
* OOR/OFFICE handling: \_\_\_\_\_\_\_\_\_\_ — Email: \_\_\_\_\_\_\_\_\_\_