Description of BioPortal First Data Release and Access Procedures
2026/02/26

This document contains a brief description of the BioPortal data that will be released during this first data release. At the end of the document are the details for access to this data

---

The BioPortal has collected samples and/or data from over 5985 individuals. This data-release contains data on up to 5985 individuals for whom chart review is complete. (Note that follow-up data entry is on-going). The characteristics of this data and sample collection are below. Here we describe the data that will be released in the first data release for use by approved investigators.

The BioPortal has recruited individuals through its own BioPortal program, which was approved by the research ethics board of the Jewish General Hospital of McGill University.

Note that this release of data represents a proportion of the data that will be released. Future data releases will require subsequent applications for data. Each application for data will require a new data transfer agreement and fee. New ethics applications will not be required if you are continuing the same research program.

Note also, that this data release is governed by the data release mechanism approved by the JGH Research Ethics Board and is not a data release organized by the BioPortal itself. The proteomics that was generated for this data arose from the local collection of DNA and plasma for the BioPortal. This data type was paid for with funds from the Roche, cqdm and JGH Foundation.
Proteomic data for this program was generated by Olink using their Explore HT Platform.
This data release process will be managed by the BioPortal Data Access Committee.

Applications for access to data are encouraged from investigators at academic institutions, within industry and from all countries, except for investigators affiliated with the named research organizations of concern for Canada: https://science.gc.ca/site/science/en/safeguarding-your-research/guidelines-and-tools-implement-research-security/sensitive-technology-research-and-affiliations-concern/named-research-organizations.

H-413, 3755, CH. DE LA CÔTE STE-CATHERINE ROAD, MONTRÉAL (QUÉBEC).
TÉL / TEL : 514-340-8222, POSTE / LOCAL : 27090.

Description of BioPortal data.

The data that will be released for this First Data Release:

1- There are 7 data sets (ds) with demographic and **clinical data**:

- ds_baseline.csv: data on baseline information collected at informed consent, such as demographic and clinical data and information on collected urine and blood samples, including availability of proteomics data.
- ds_medical_history.csv: repeating measures on medical history of patients and their relatives.
- ds_laboratories.csv: repeating laboratory measures.
- ds_icd10.csv: repeating measures on diagnosis coded by international classification of disease version (ICD-10 English).
- ds_bioimpedance.csv: repeated measures on body composition and mass
- ds_cardiovascular.csv: repeated measures on pulse and blood pressure.
- ds_retinopathy.csv: repeated measures on diabetic retinopathy and macular edema.

2- There are 2 datasets of **proteomics data** along with 11 PDFs documenting quality checks and data descriptions.

- ANNOTATED_ DEIDENTIFIED_Release1_Proteomics_Subset.parquet: Proteomic dataset post quality checks with additional annotation columns for initial proteomic data release.
- DEIDENTIFIED_Release1_Proteomics_Subset.parquet: Deidentified raw proteomics data for initial proteomic data release.
- step_{0 – 10}*.pdf: which contains description of quality checks, source code, and data column descriptions.

Clinical data is detailed in the Data Dictionary entitled:
data_dictionaries.pdf

The non-genomic data will be shared in a ".csv "format. We strongly suggest not to use Excel to view or manipulate this data.
The proteomics data will be shared in a ".parquet "format. Note that due to differences in meta-data information, this file cannot be opened directly with OlinkAnalyze::read_NPX().

H-413, 3755, CH. DE LA CÔTE STE-CATHERINE ROAD, MONTRÉAL (QUÉBEC).
TÉL / TEL : 514-340-8222, POSTE / LOCAL : 27090.

Brief Description of Proteomics Data

1. Data was generated using the Olink Explore HT platform.
2. This included 5,416 assays targeting 5,416 proteins.
3. Sample linking file contains a random sample ID and the corresponding project ID unique to each project.
    - Note, sample ID within the proteomic data files is of the format <PlateID>_<random Sample ID>. To obtain the corresponding project ID with the linking file, you must first remove the PlateID information.
4. Details of the protein targeted can be found on the Olink webpage.
5. Details of protein quality control checks and source code can be found in the accompanying documentations.
6. Two datasets are provided:
    - Raw results
    - Results post quality checks with additional annotation columns

~~H~~-413, 3755, CH. DE LA CÔTE STE-CATHERINE ROAD, MONTRÉAL (QUÉBEC).
TÉL / TEL : 514-340-8222, POSTE / LOCAL : 27090.

Data types in BioPortal

Clinical data

- Demographic information
- International Classification of Disease Version 10 (ICD-10)
- Medical history of patient and family
- Diastolic and systolic blood pressure and heart rates
- Chronic kidney disease staging (CKD-EPI 2021)

Retinal scans

- Full retinal images (both eyes)
- Diabetic macular edema detection
- Diabetic retinopathy severity score

Laboratory

- Albumin
- ALT
- Cholesterol total
- Creatinine (random urine)
- Glucose
- HDL-Cholesterol
- IGF1
- LDL-Cholesterol
- Lipoprotein(a)
- Microalb. to creatinine ratio
- Microalbumin (random)
- Total arterial hemoglobin
- Total venous hemoglobin
- Triglycerides
- Urea
- …multiple other values

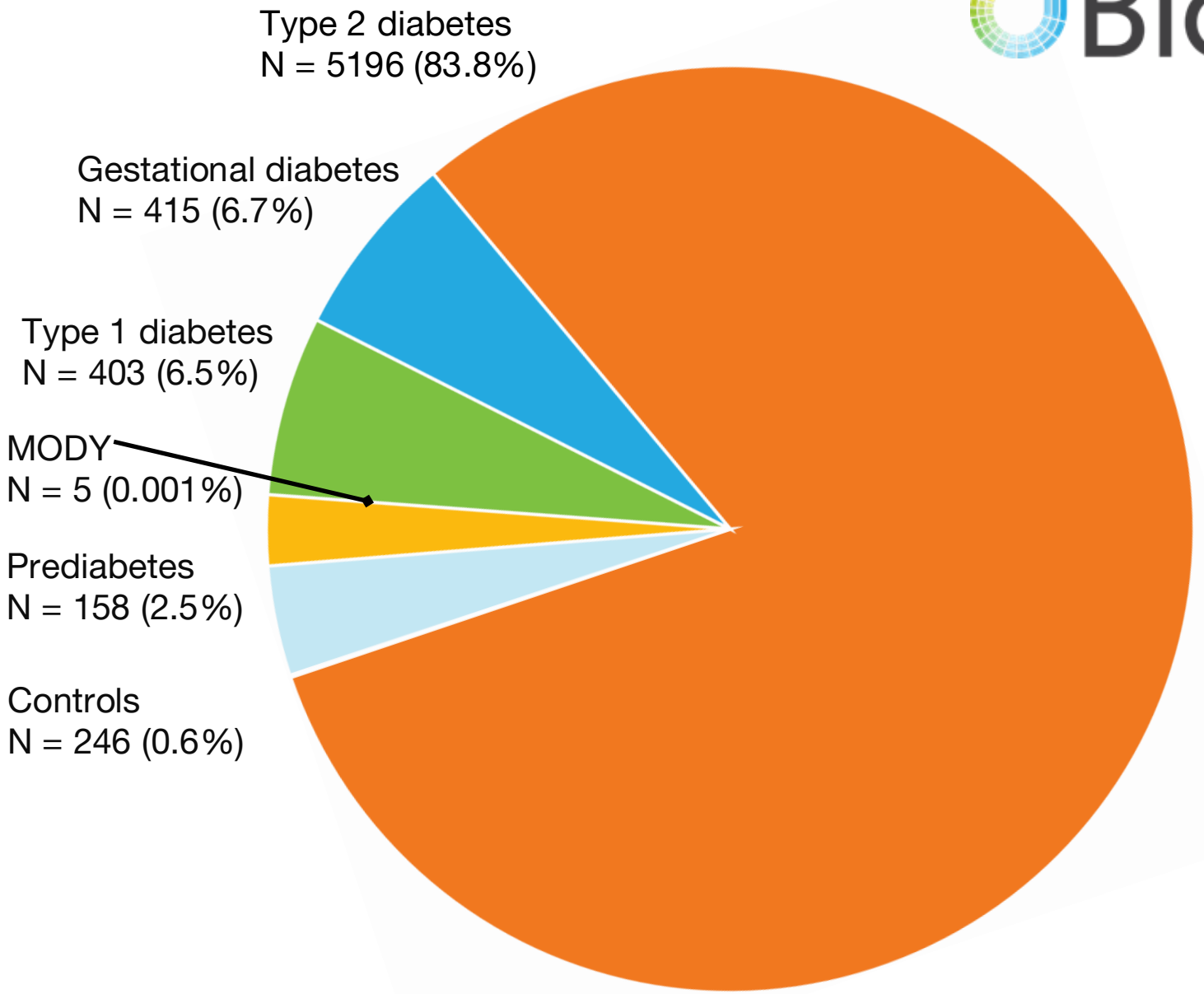
Biobank

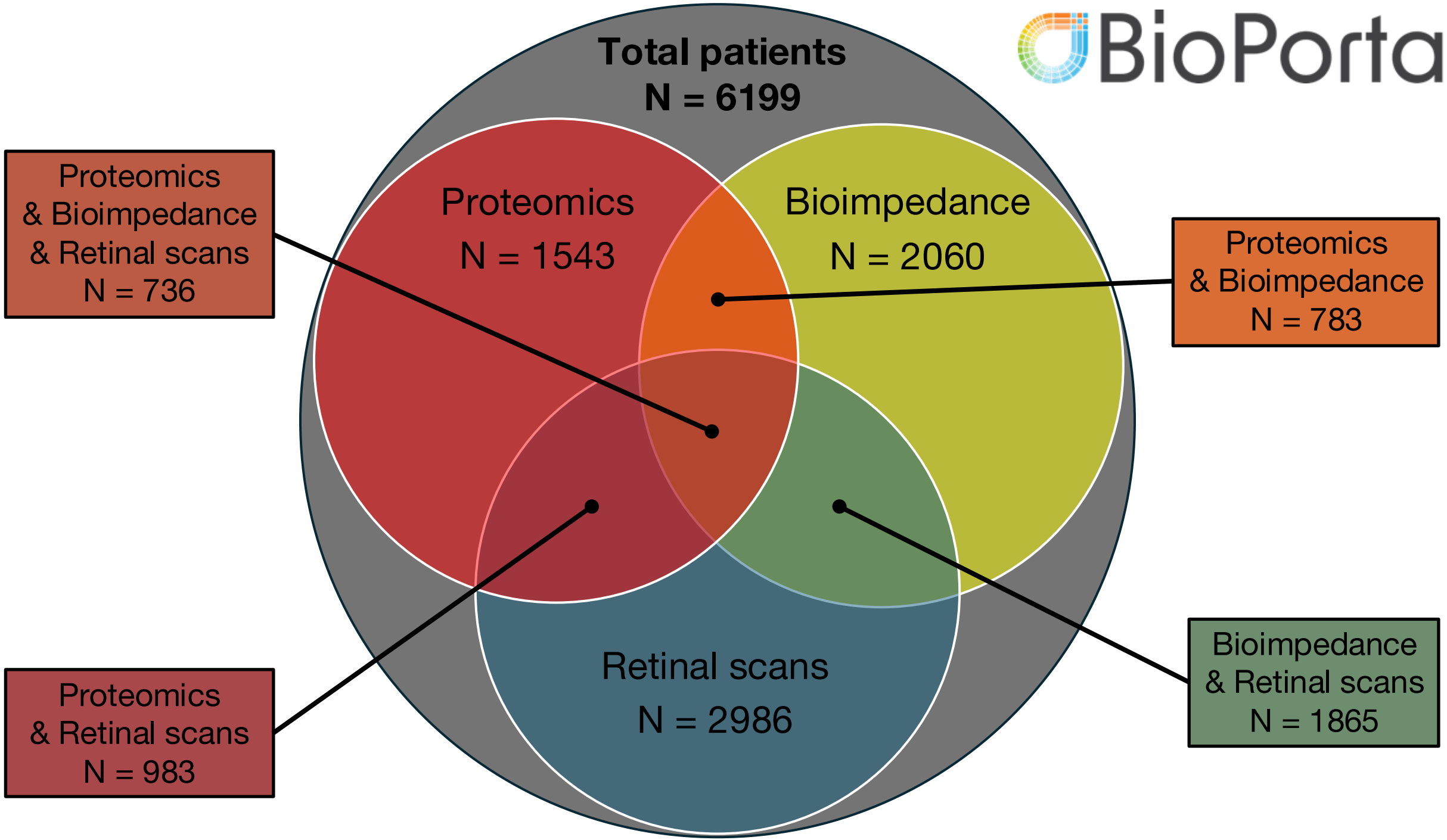
- Blood samples
- Saliva
- Urine samples

Bioimpedance

- Fat mass
- Bone mass
- Muscle mass
- BMI

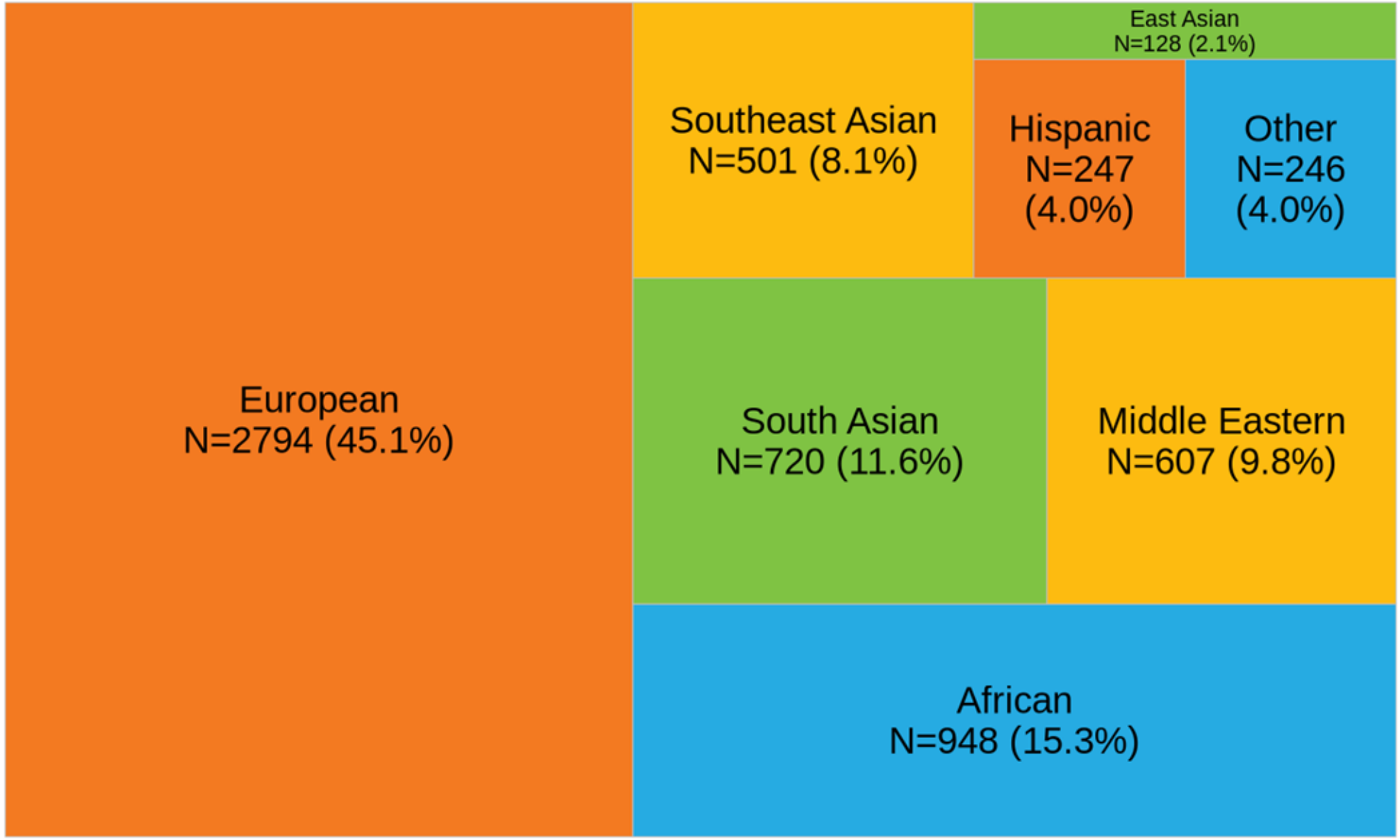
Total patients
in BioPortal
N = 6199



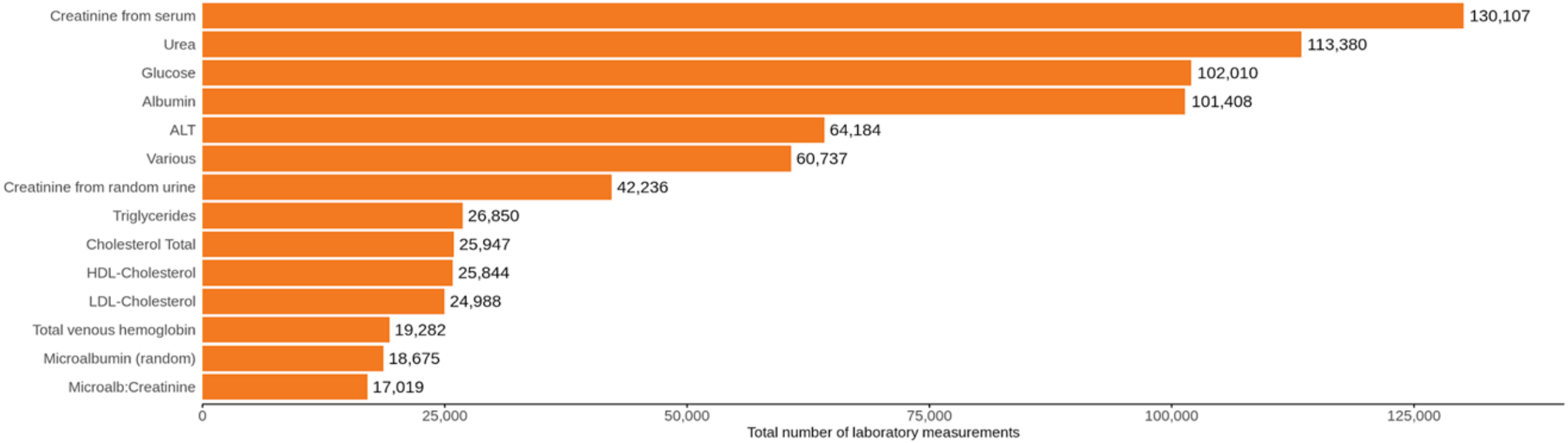

Total patients
N = 6199
Proteomics
N = 1543
Bioimpedance
N = 2060
Retinal scans
N = 2986
Proteomics
& Bioimpedance
& Retinal scans
N = 736
Proteomics
& Bioimpedance
N = 783
Proteomics
& Retinal scans
N = 983
Bioimpedance
& Retinal scans
N = 1865

Self-reported ethnicity

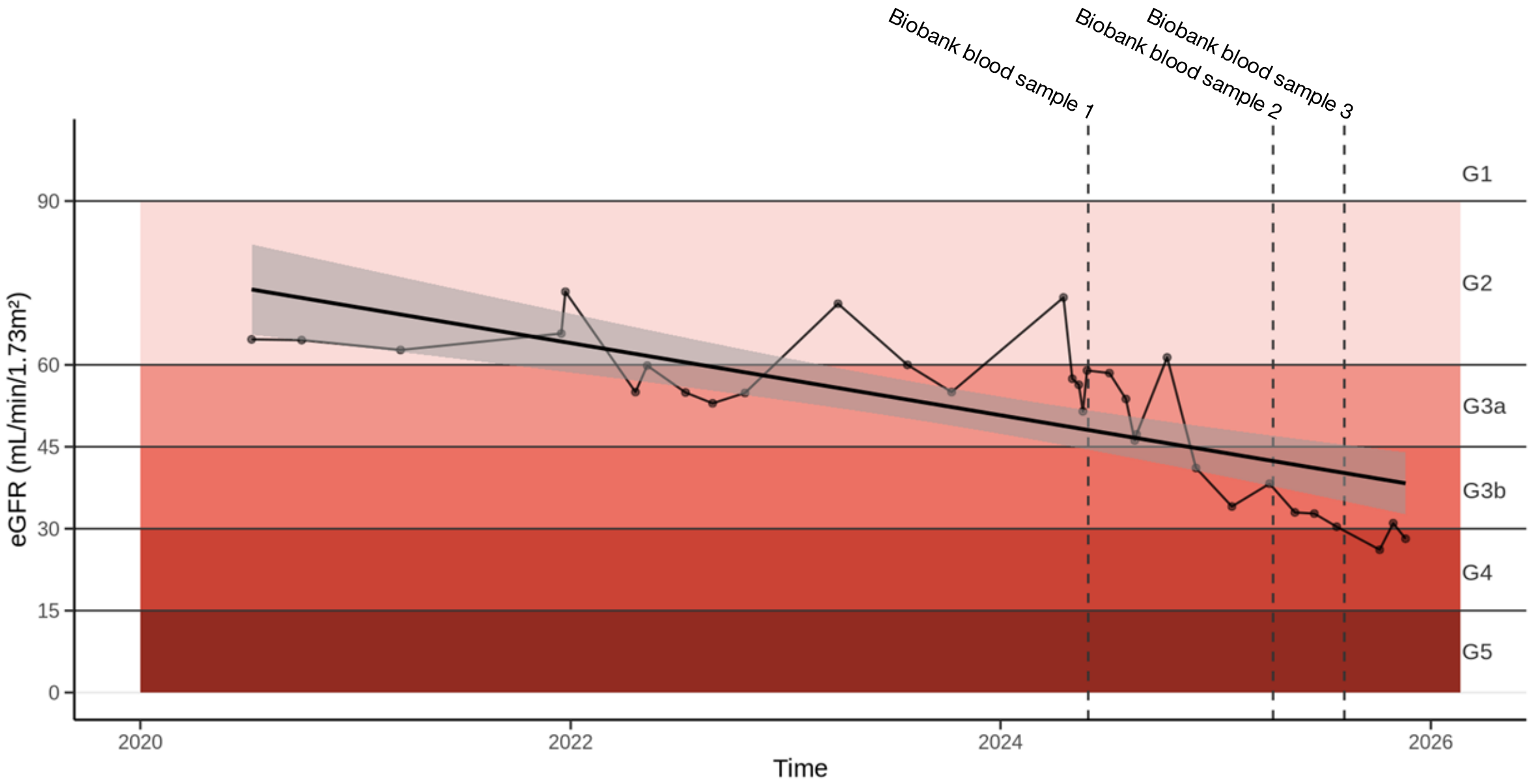
BioPortal



Laboratory measurements from clinical routine



eGFR progression in single patient



Process for Data Access:
All applicants must fill out the BioPortal data access application form, with an accompanying one-page project description. This must be submitted to access.bioportal@ladydavis.ca

Applications will be judged to ensure that they ask a scientific or medical question and that the applicant agrees to respect the data confidentiality and other conditions below.

1. The proposed project must be subject to approval and oversight by the applicant's research ethics board (REB). The applicant must provide written approval of the project by their REB, for secondary use of data. This proof does not need to accompany the data access application form, but no data will be released without REB approval.
2. Data Access Pricing:
    a. Base Pricing (Academic Rate)

Data / Service	Academic Price	Approx. Sample Size	Notes
Clinical data	$3,500	~6,000 individuals	—
Genotype data	$4,000	~6,000 individuals	—
Olink proteomics	$7,000	1,543 individuals	—
Professional services*	$190 / hour	—	Imaging dataset preparation

b. Business Pricing Multipliers

Organization Type	Definition	Pricing Multiplier	Example (Clinical Data)
Small & Medium Businesses	< 100 employees	1.5× academic rate	$5250
Large Businesses	> 100 employees	4.0× academic rate	$14000

c. Introductory discount policy: discount eligibility is determined based on the date of the completed application submission.

Eligible Category	Discount	Duration / Limit
Academics	50% off	First 3 months after announcement
Small & Medium Businesses	50% off	First 3 months after announcement
Large Businesses	50% off	First 3 large business customers

d. Additional Notes:
   i. $500 administrative fee per application
   ii. All prices are listed in CAD.
   iii. Prices do not include dataset updates.
   iv. We are accepting custom data requests. Depending on the scope, these may incur professional service fees* for data preparation. For example, imaging data can be provided on a request basis, and all images undergo metadata de-identification prior to release.

3. Applicants should request the genetic data only if they have approval from their REB to undertake genetic analyses.

4. The applicant must enter into a data access agreement (DTA) with the JGH. In this DTA, the following stipulations will be made:

a. The applicant must not share any data with any investigators not listed in the application.
b. The applicant must agree to acknowledge that data was provided by the "BioPortal". No authorship is requested by the BioPortal.
c. The applicant must agree to return any data derived from analysis of BioPortal data to the BioPortal for broad sharing, at the time that you submit your derived data for peer-review
d. The applicant must agree not to attempt to re-identify any individual within the BioPortal.
e. The applicant must agree to safeguard the data using the below safeguards as a minimum:

i. All computers with access to the Information must employ logical access controls (passwords) at the device and network level.
ii. Where the Information is held on laptops, CD-ROMs, flash memory sticks or other transportable media of any type, passwords and full encryption must be used. This applies equally to backups of the Information stored on transportable media.
iii. The Information cannot be electronically transmitted, except as described below. This includes the transmittal of the Information by facsimile or by e-mail.
iv. Servers storing and transmitting unencrypted data, where used, must be located in a secure, controlled-access area, preferably in the same area where the Information is accessed. If located in a separate area, controls must be in place to ensure that only approved researchers can access the server. Unless the Information is encrypted continuously while outside the secure area, conduit must be used for all cabling and all cross-connect areas must be physically secured.
v. Network firewalls and access rules must be in place to prevent access to the Information, other than to approved researchers. Information may be stored on and transmitted over networks not meeting these requirements, provided that it is encrypted, except when in use by an Identified Person. Alternatively, the Information may be stored on a stand-alone computer with no external connections, or on a closed network. When a network transmits information that leaves a secure area (for example, when a series of buildings house employees within a single organization), the data must be encrypted whenever it is outside the secure area.