VANDERBILT UNIVERSITY
MEDICAL CENTER

Information in this presentation is current as of 10/27/2014

BioVU and the Synthetic Derivative

Sarah Collier, PhD

Project Manager, BioVU

The Take Home Message

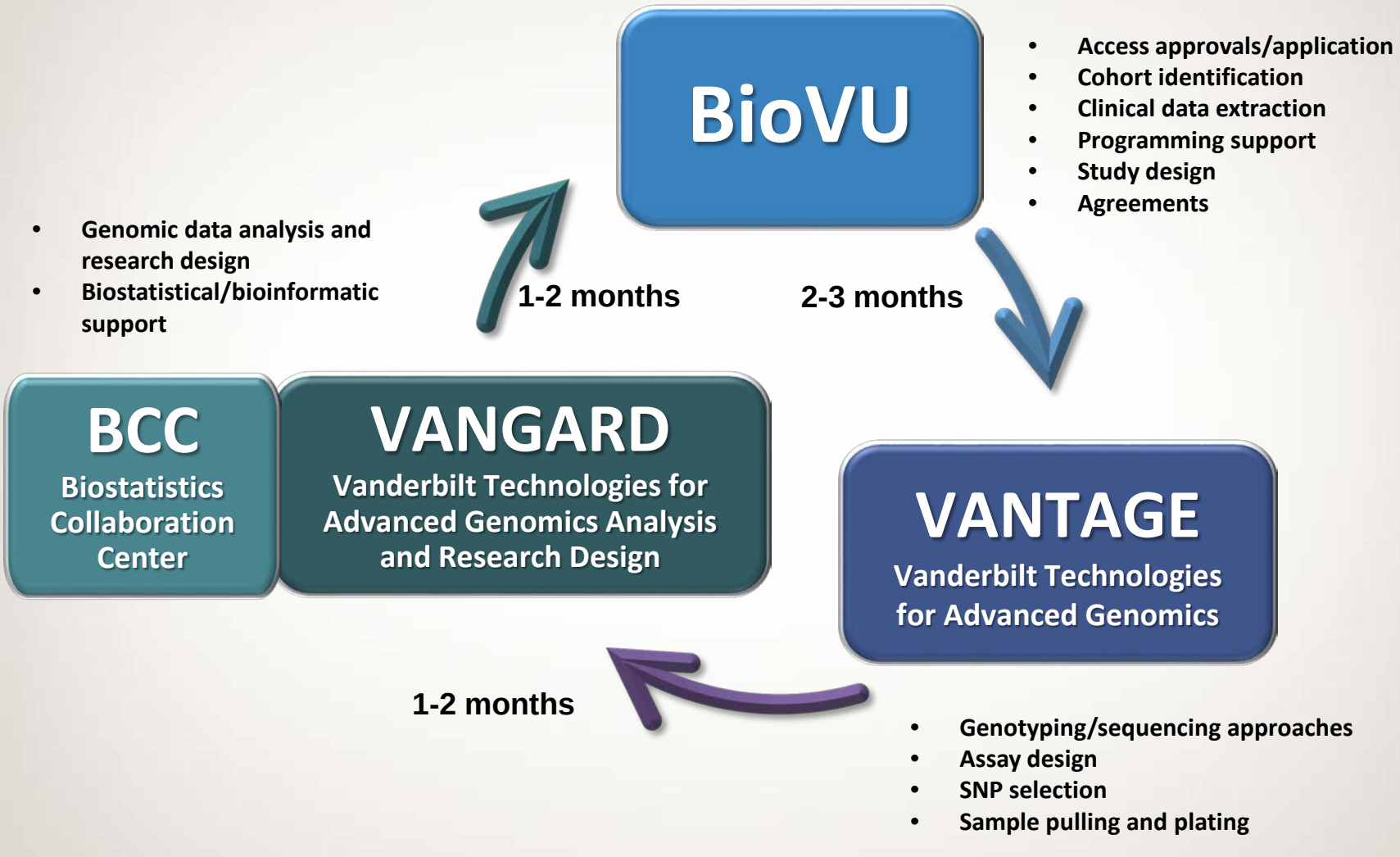
BioVU and the Synthetic Derivative Resources are:

- Readily available for use
- Broad utility
- Cost effective
- Efficient

Resources and Tools are available at all stages!

VANDERBILT UNIVERSITY
MEDICAL CENTER

BioVU Project Life Cycle

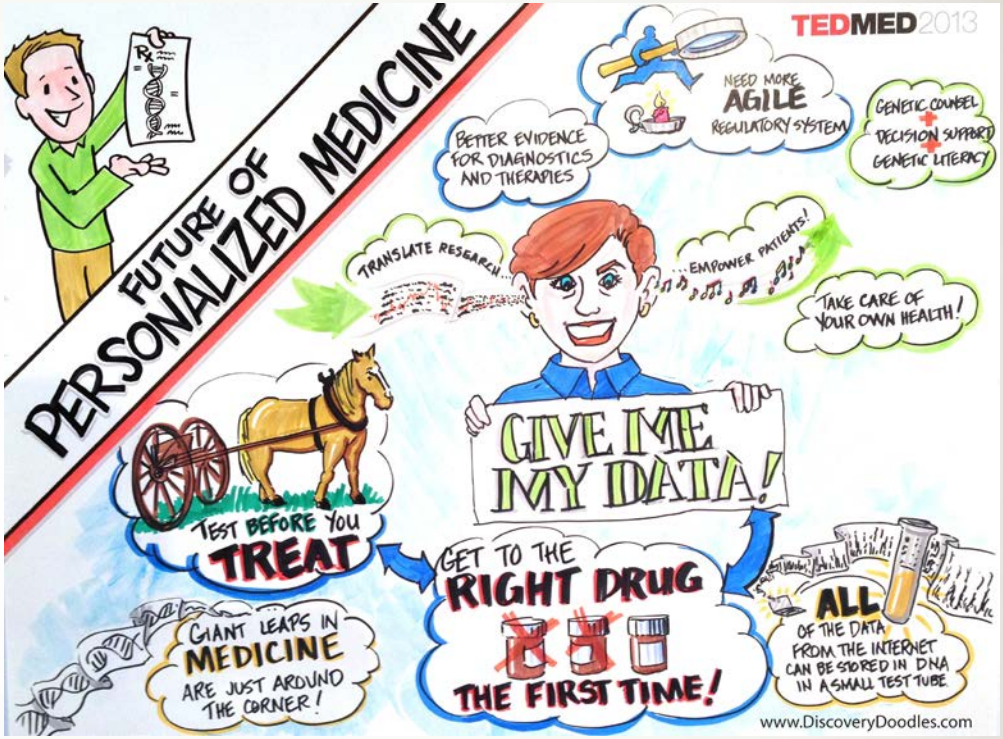


VANDERBILT UNIVERSITY
MEDICAL CENTER

Personalized Medicine

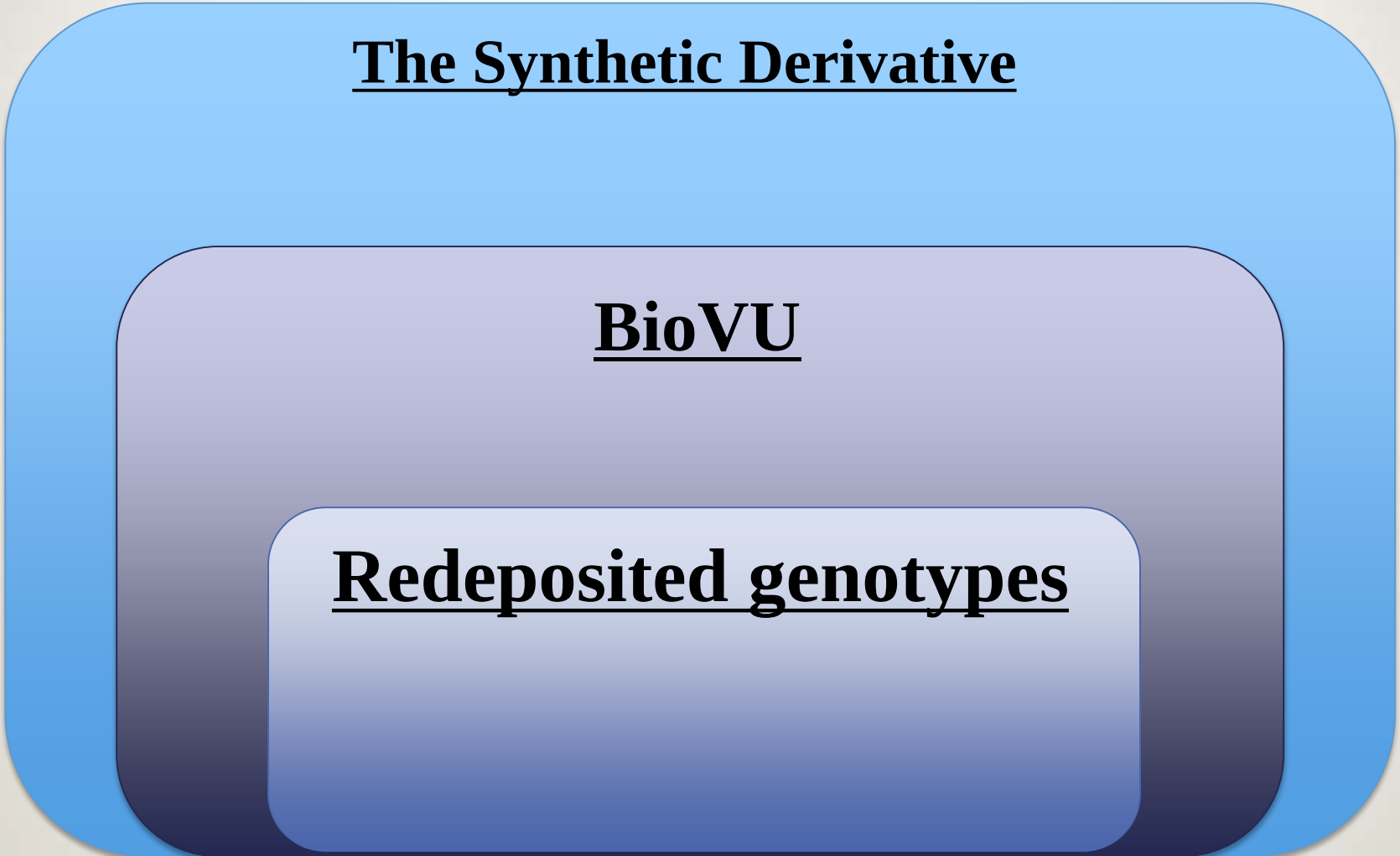
- Shift the emphasis in medicine from reaction to prevention
- Direct the selection of optimal therapy and reduce trial-and-error prescribing
- Help avoid adverse drug reactions
- Increase patient adherence to treatment
- Improve quality of life
- Reveal additional or alternative uses for medicines and drug candidates
- Help control the overall cost of health care

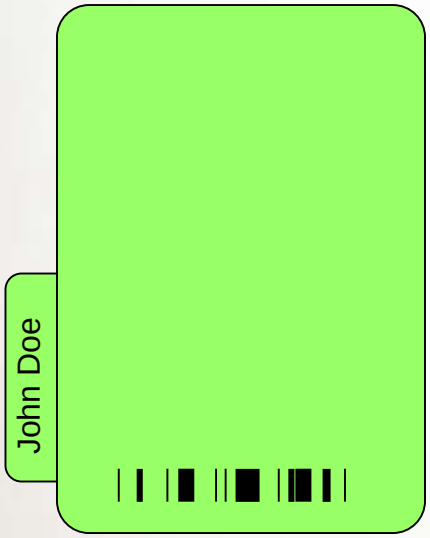
The Case for Personalized Medicine from the Personalized Medicine Coalition 4th Edition 2014.

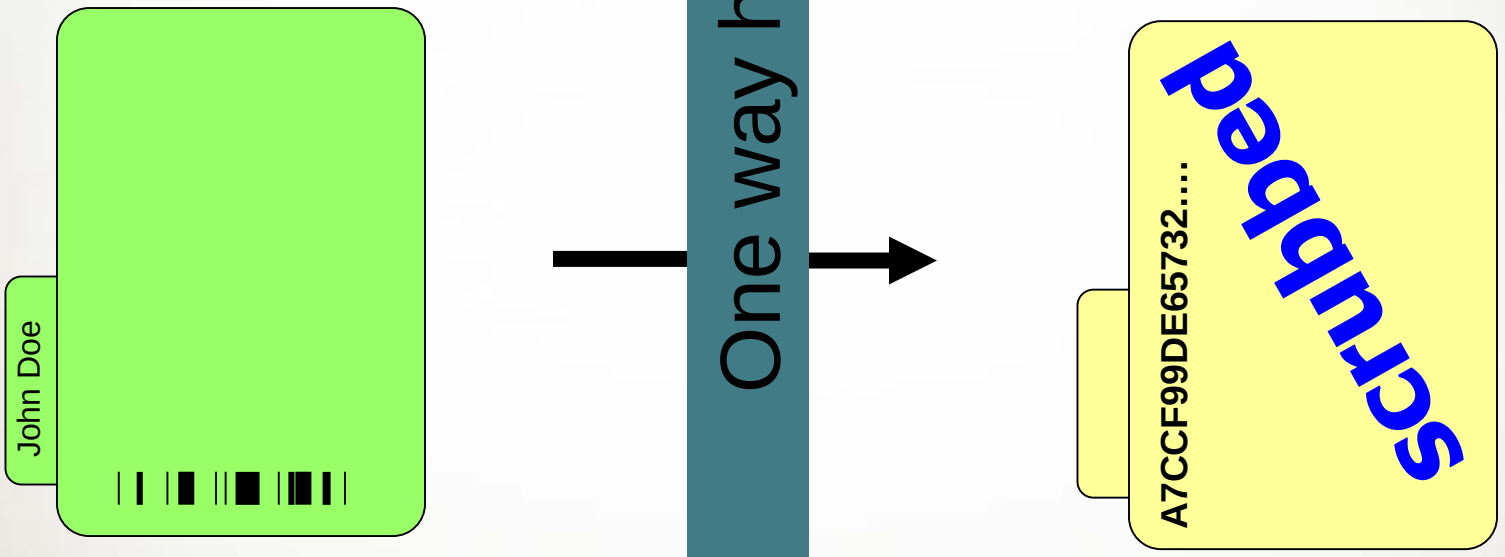


VANDERBILT UNIVERSITY
MEDICAL CENTER

Resources for EMR-based research at VUMC



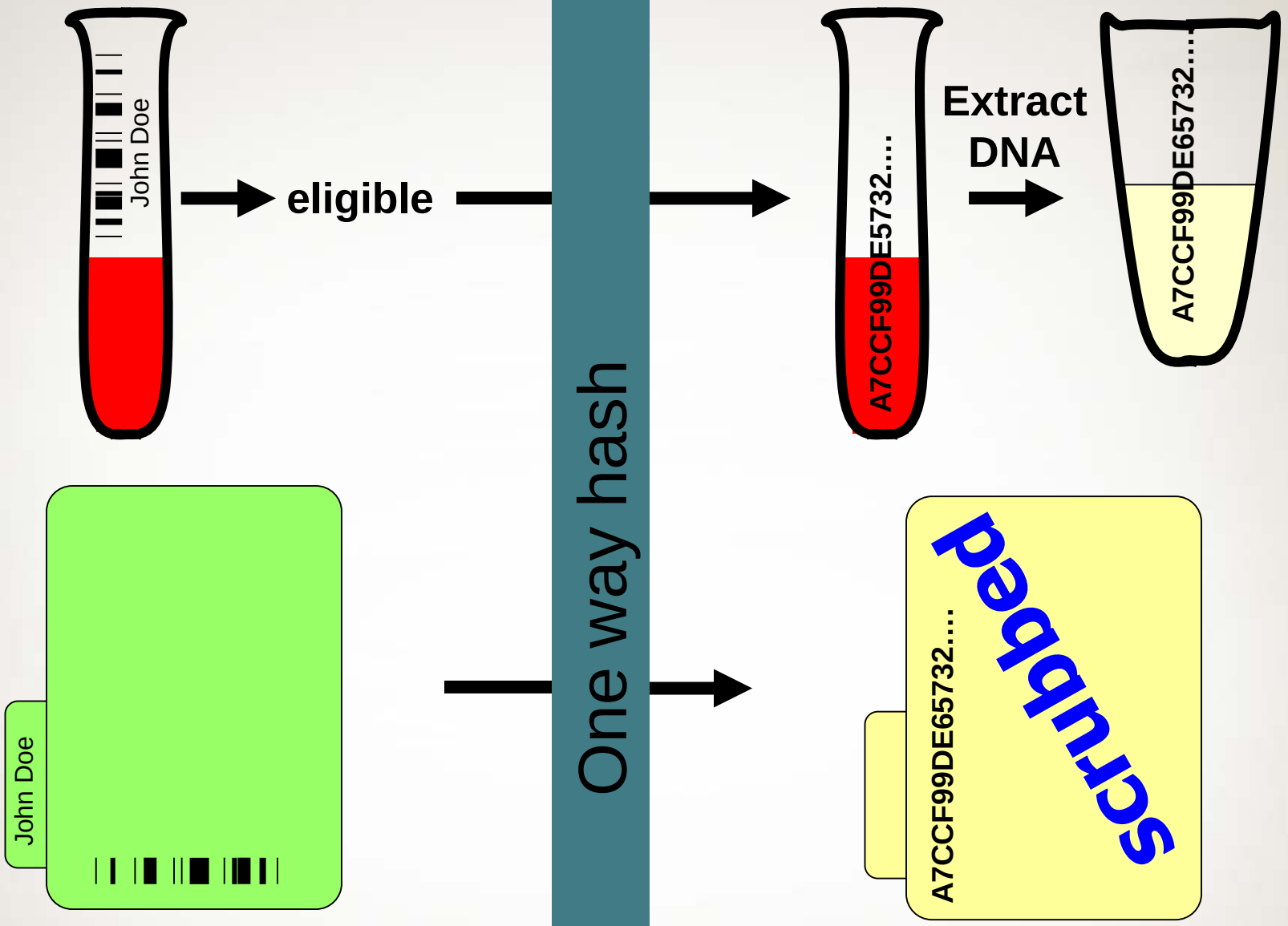

John Doe



>2 million records

The Synthetic Derivative:
can be updated

VANDERBILT UNIVERSITY
MEDICAL CENTER



>2 million records

The Synthetic Derivative:
can be updated

VANDERBILT UNIVERSITY
MEDICAL CENTER

VANDERBILT UNIVERSITY
MEDICAL CENTER

Resources for EMR-based research at VUMC

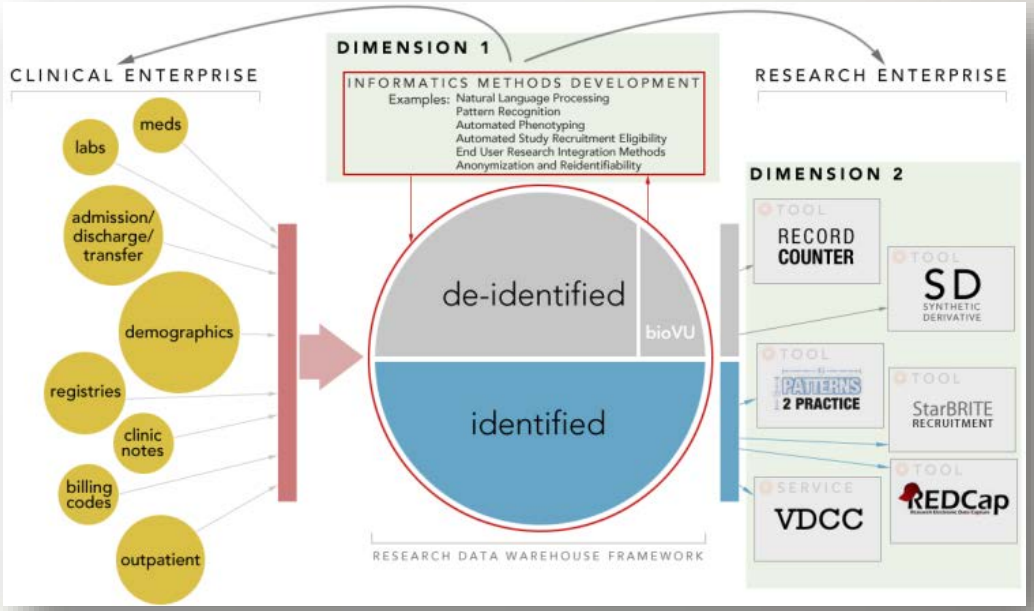
The Synthetic Derivative

A de-identified and continuously-updated image of the EMR (>2 M records)

VANDERBILT UNIVERSITY
MEDICAL CENTER

Synthetic Derivative

- A **Derivative** of the EMR - information content reduced by 'scrubbing' identifiers
- User Interface tool that can be used for access and analysis
- Services are available to help deliver results for non-standard queries (temporal queries, controls matching, etc)
- Contains >2.0 million records
  - ~1 million with detailed longitudinal data
  - averaging 100,000 bytes in size
  - an average of 27 codes per record
- Records updated over time and are current through 07/2014



VANDERBILT UNIVERSITY MEDICAL CENTER

Synthetic Derivative Data Types

- Narratives, such as:
  - Clinical Notes
  - Discharge Summaries
  - History and Physicals
  - Problem Lists
  - Surgical Reports
  - Progress Notes
  - Letters
- Diagnostic Codes, Procedural Codes
- Forms (intake, assessment)
- Reports (pathology, ECGs, echocardiograms)
- Clinical Communications
- Lab Values and Vital Signs
- Medication Orders
- TraceMaster (ECGs)
- Tumor Registry

VANDERBILT UNIVERSITY
MEDICAL CENTER

Technology + Policy

De-identification

- Derivation of 128-character identifier (RUI) from the MRN generated by Secure Hash Algorithm (SHA-512)
- HIPAA identifiers removed using combination of custom techniques and established de-identification software

Date Shift

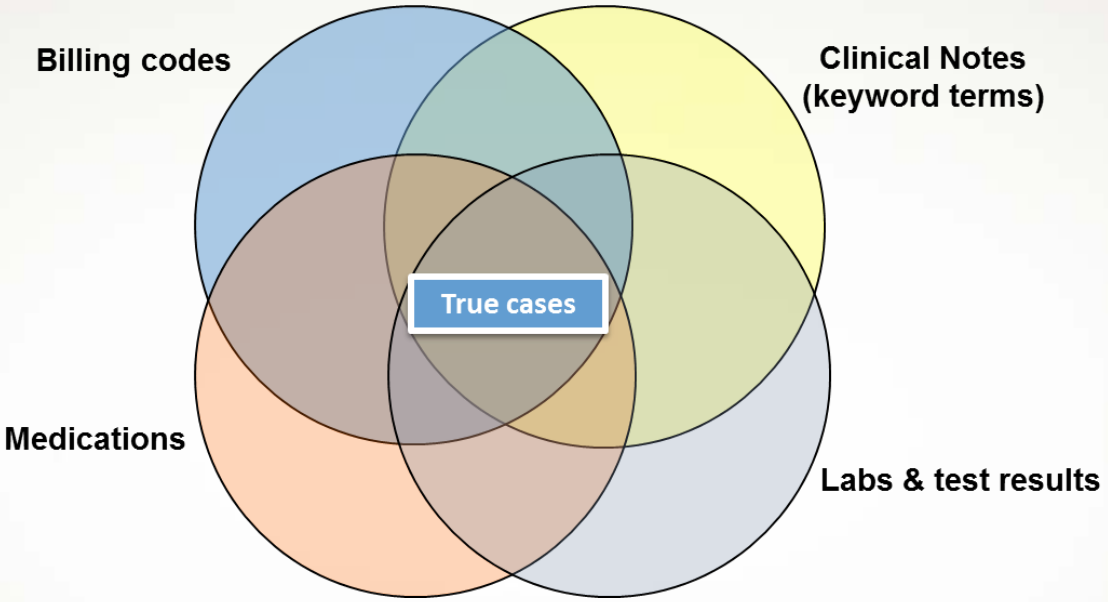
- Our algorithm shifts the dates within a record by a time period *(up to 364 days backwards)* that is consistent within each record, but differs *across* records

Restricted access & continuous oversight

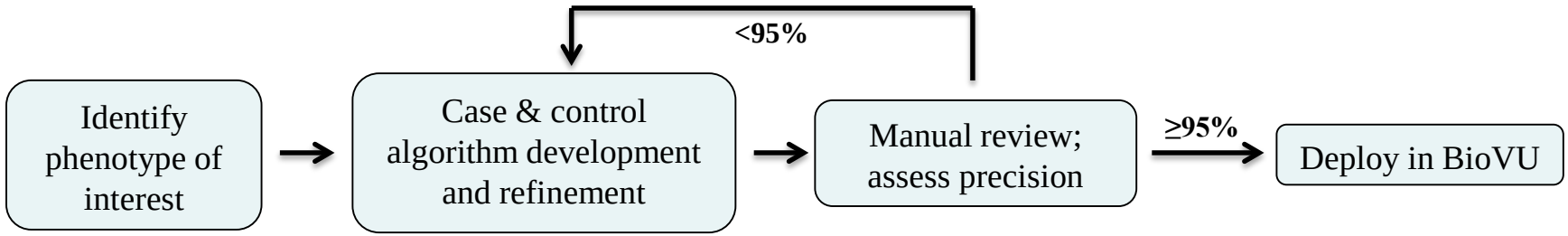
- Access restricted to VU; not a public resource
- IRB approval for study (non-human)
- Data Use Agreement
- Audit logs of all searches and data exports

VANDERBILT UNIVERSITY
MEDICAL CENTER

Phenotyping Approach

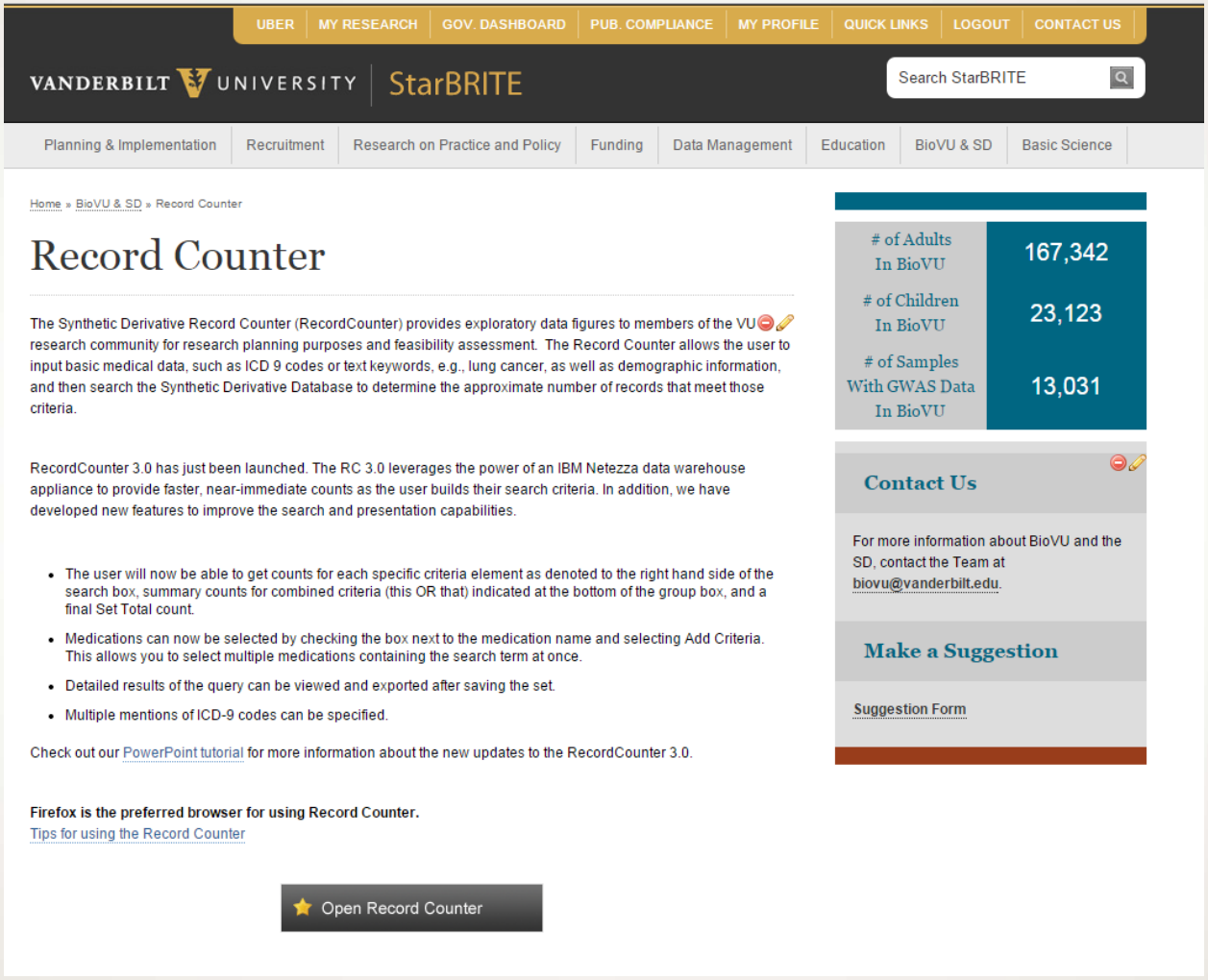


Algorithm Development



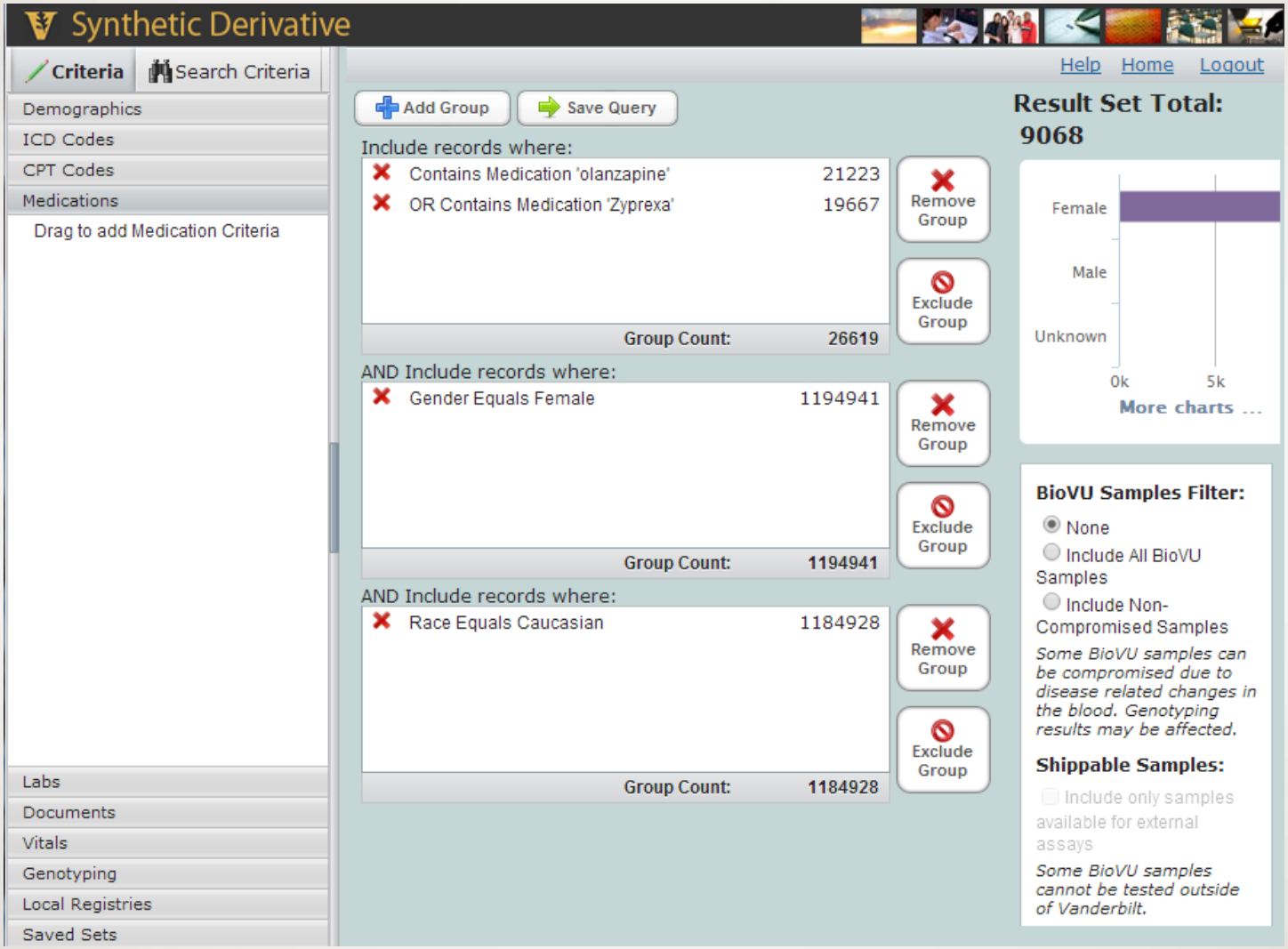
VANDERBILT UNIVERSITY MEDICAL CENTER

Feasibility Counts Record Counter Tool



VANDERBILT UNIVERSITY
MEDICAL CENTER

Synthetic Derivative Use



VANDERBILT UNIVERSITY
MEDICAL CENTER

Synthetic Derivative Access

Planning & Implementation | Recruitment | Research on Practice and Policy | Funding | Data Management | Education | BioVU & SD | Basic Science

Home » BioVU & SD » Application Instructions » Synthetic Derivative Access Only

Synthetic Derivative Access Only

Components to Access the Synthetic Derivative

For Primary Investigators:

1. **Download the SD Access Application Checklist**
2. **Obtain a Non-Human Subjects Determination Letter from the IRB:**

Contact Us

For more information about BioVU and the SD, contact the Team at biovu@vanderbilt.edu.

Make a Suggestion

Suggestion Form

VANDERBILT UNIVERSITY
MEDICAL CENTER

Resources for EMR-based research at VUMC

The Synthetic Derivative

A de-identified and continuously-updated image of the EMR (>2 M records)

BioVU

- DNA samples available: >188,000
- Plasma collection underway

VANDERBILT UNIVERSITY
MEDICAL CENTER

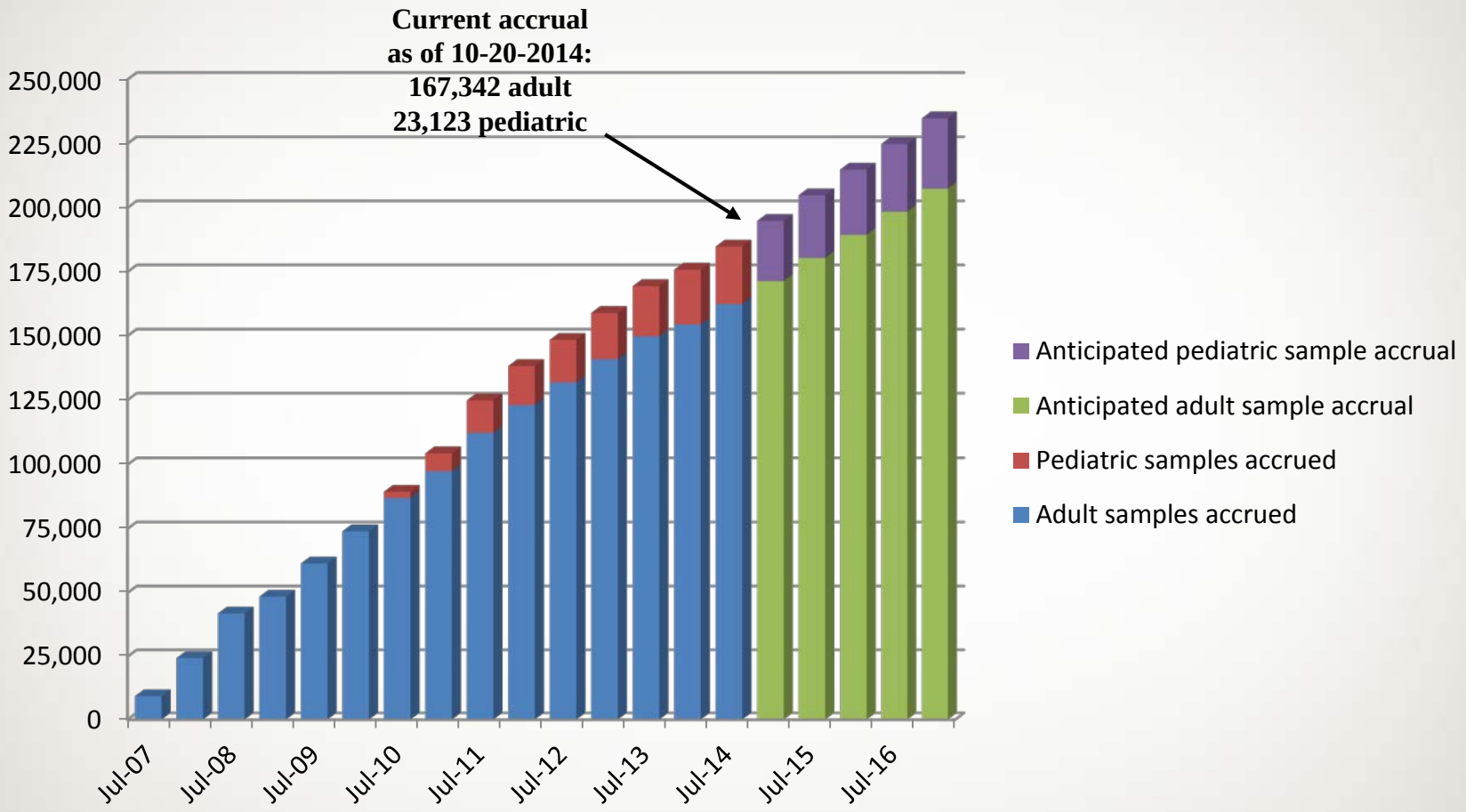
BioVU Sample Acceptance

Accepted samples must:

- Be of good quality
- Have sufficient amount of blood
- Be from a patient who has signed the BioVU form
- Be from a patient who has not opted out

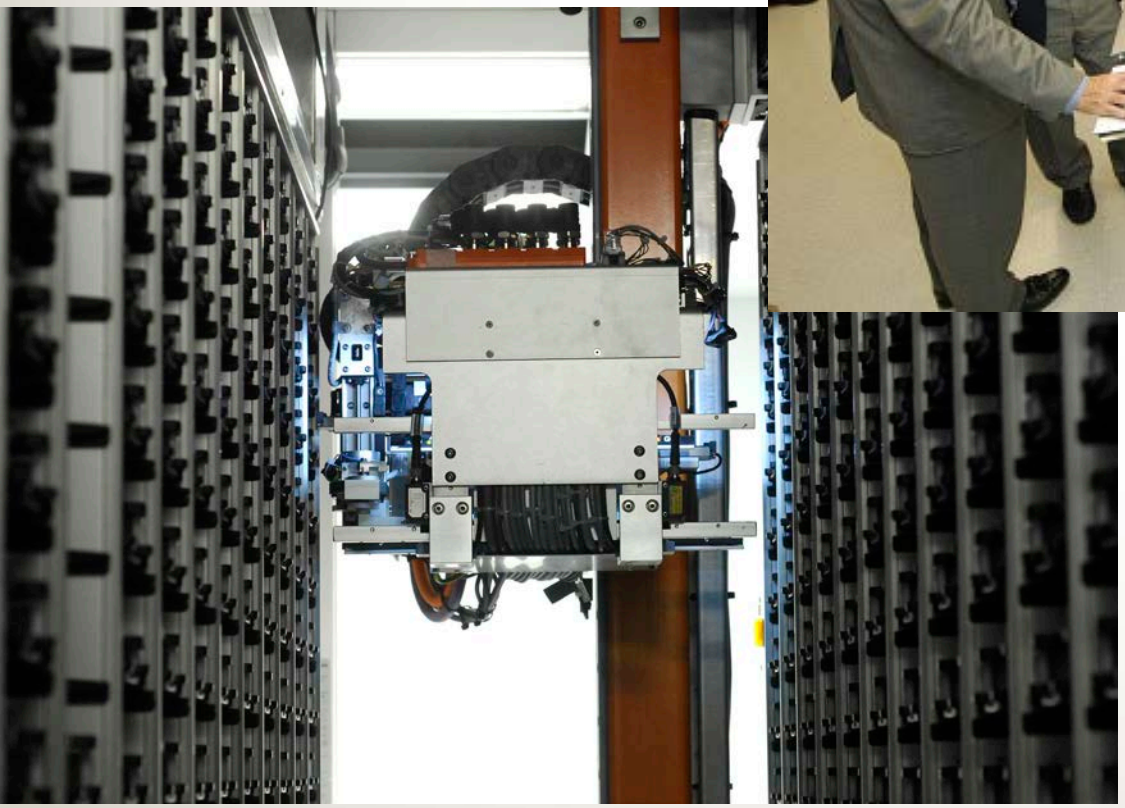
VANDERBILT UNIVERSITY
MEDICAL CENTER

BioVU Sample Accrual: 190,336



VANDERBILT UNIVERSITY
MEDICAL CENTER

BioVU Storage



RTS SmaRTStore

VANDERBILT UNIVERSITY
MEDICAL CENTER

The BioVU Form

A component of the Consent for Treatment process

DNA Research:

Leftover blood from tests, treatment, or surgery may also be used for DNA research through the Vanderbilt **BioVU** Program. If I do not want my leftover blood to go to the Vanderbilt BioVU Program for DNA research, I must check the box below. If I have questions or want further information on **BioVU**, I may call 866-436-4710.

☐ I do **NOT** want blood left over from my tests, treatment, or surgery to be used for the Vanderbilt BioVU Program for DNA research.

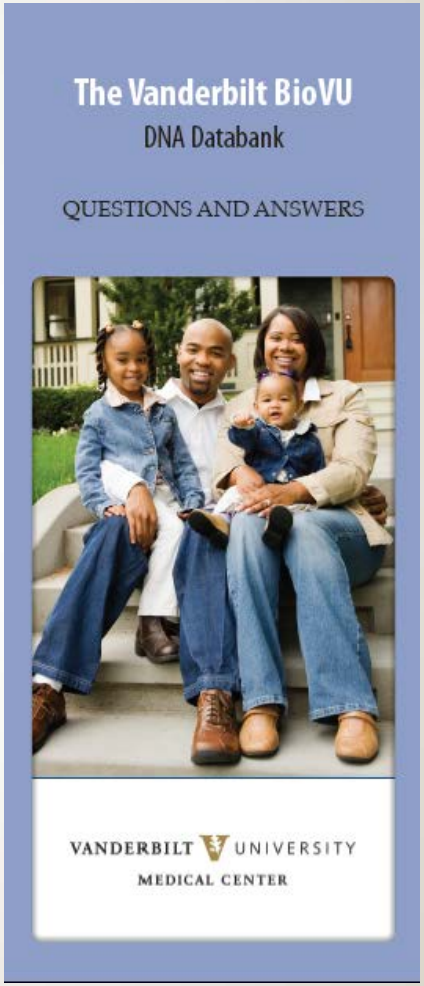
Patient/Legal Representative _____ Date _____

(Relationship to Patient) _____

VANDERBILT UNIVERSITY
MEDICAL CENTER

Patient and Community Awareness

- Posters in phlebotomy areas in English and Spanish
- Brochures freely available to VUMC clinics in English and Spanish
- BioVU hotline available for questions and opt-out



VANDERBILT UNIVERSITY
MEDICAL CENTER

Disease Cohorts

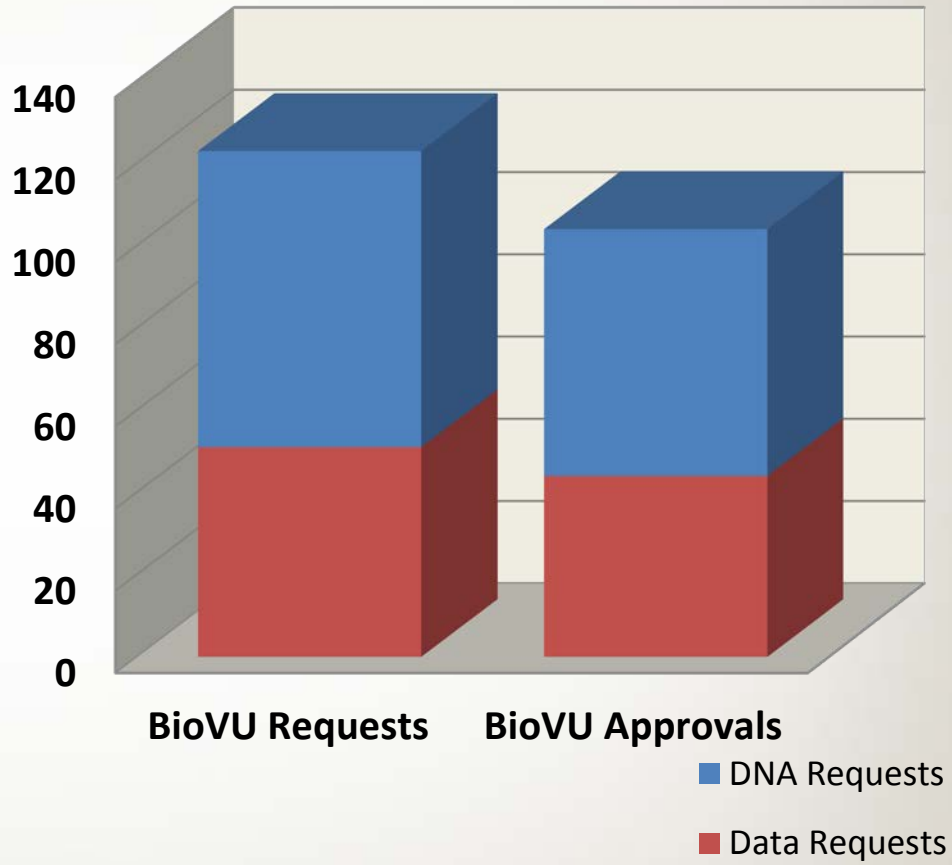
Rare Diseases

DISEASE	SD	BioVU
Pompes Disease	26	7
Duchenne / Becker Muscular Dystrophy	155	20
Glycogen storage disorders	169	45
Charcot Marie Tooth Disease	176	46
Myotonic Dystrophy	223	40
Normal Pressure Hydrocephalus	310	140
Gaucher Disease	388	115
Marfan Syndrome	812	92
Osteomyelitis	1067	154
Biliary Cirrhosis	1115	448
Myasthenia Gravis	1229	247
Microcephalus	1252	97
Systemic Sclerosis	1487	522
Polycythemia Vera	1554	451
Sarcopenia	1886	433
Sarcoidosis	2621	659
Cachexia	2787	799
Polyarteritis Nodosa And Allied Conditions	3152	698

VANDERBILT UNIVERSITY MEDICAL CENTER

BioVU Utilization

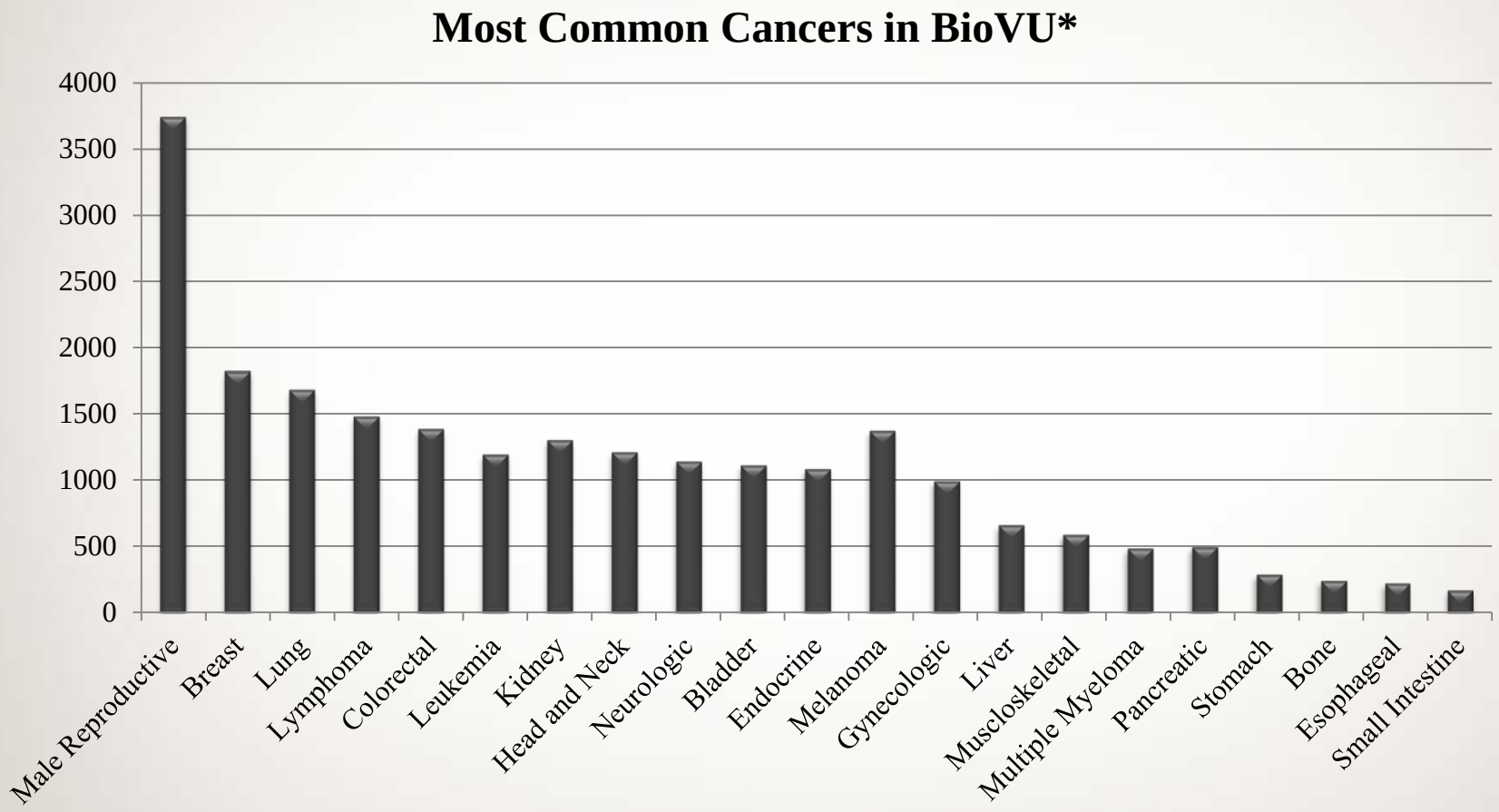
- Pre-Review
- BioVU Committee Review
  - Expedited Review*
    - Genotyping data requests
    - Reviewed by BioVU Chair
  - Full Review
    - DNA sample access requests
    - Reviewed and scored by Primary and Secondary reviewers
- BioVU Projects:
  - **Requests: 123**
  - **Approved so far: 104**



*coming Nov 1st, changes to Expedited Review

VANDERBILT UNIVERSITY
MEDICAL CENTER

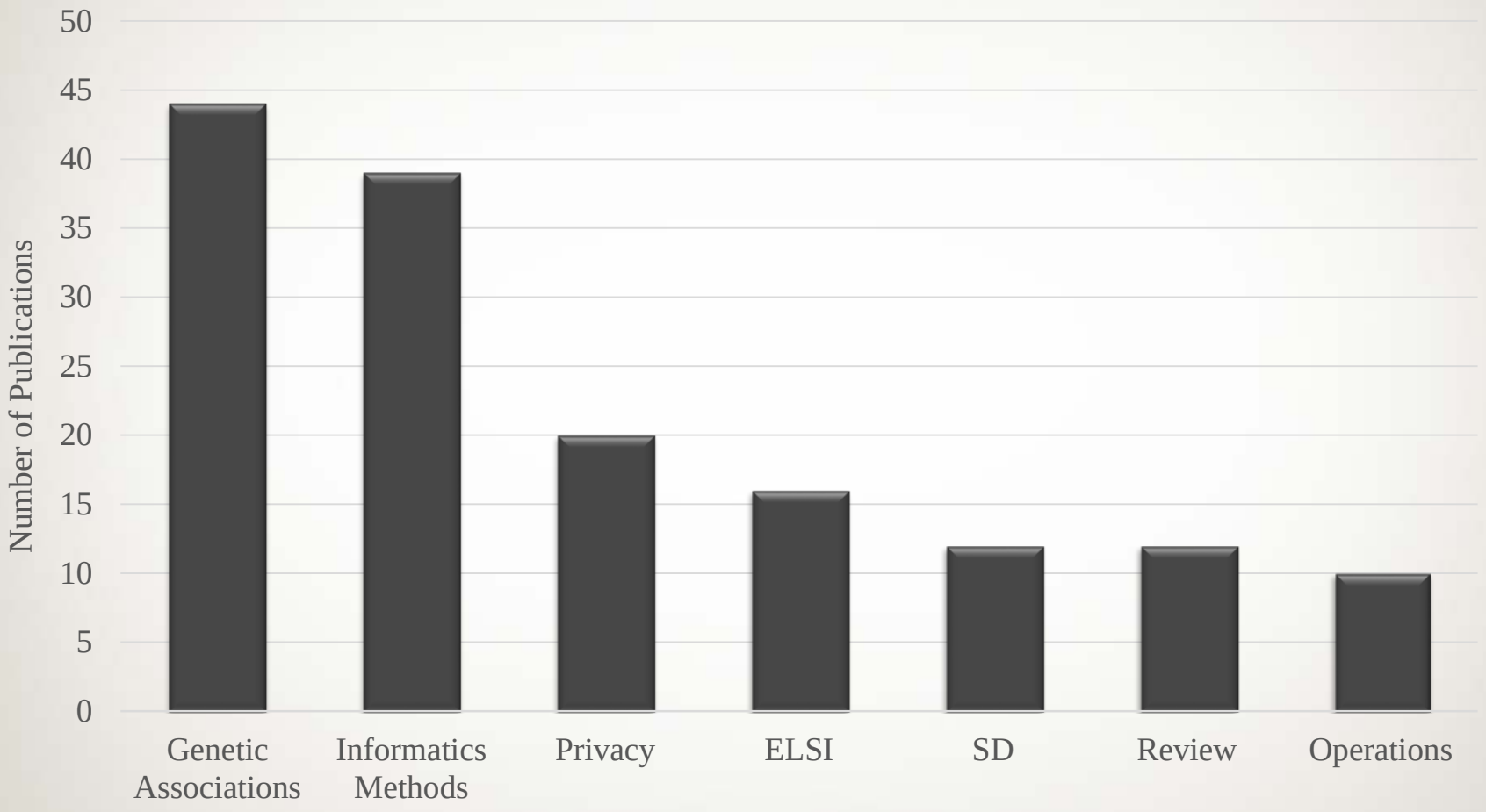
Disease Cohorts
Oncology



*as defined by Tumor Registry, September 2014

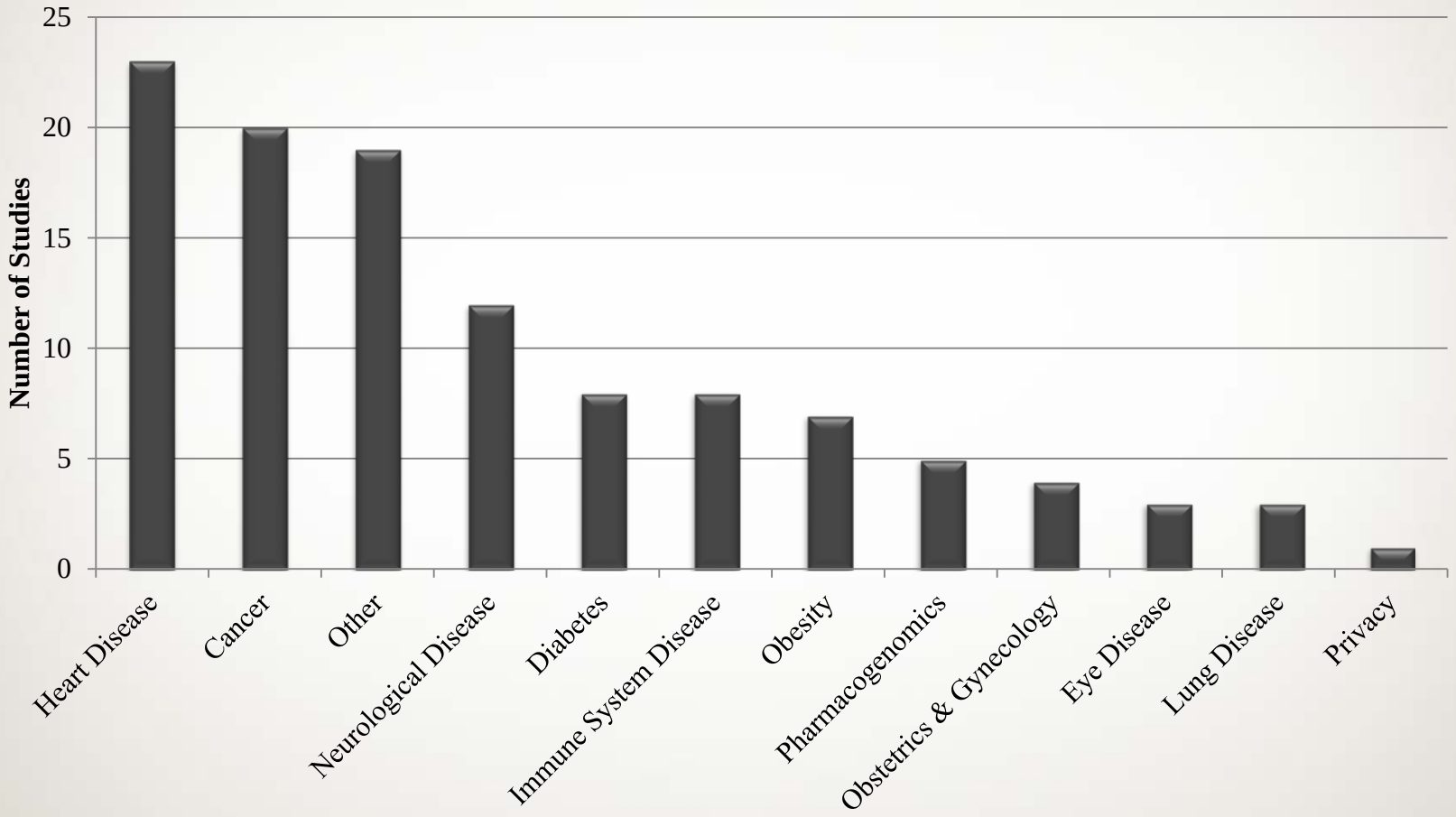
VANDERBILT UNIVERSITY
MEDICAL CENTER

BioVU and SD Publication Topics



VANDERBILT UNIVERSITY
MEDICAL CENTER

Breadth of BioVU Studies



VANDERBILT UNIVERSITY
MEDICAL CENTER

Resources for EMR-based research at VUMC

The Synthetic Derivative

A de-identified and continuously-updated image of the EMR (>2 M records)

BioVU

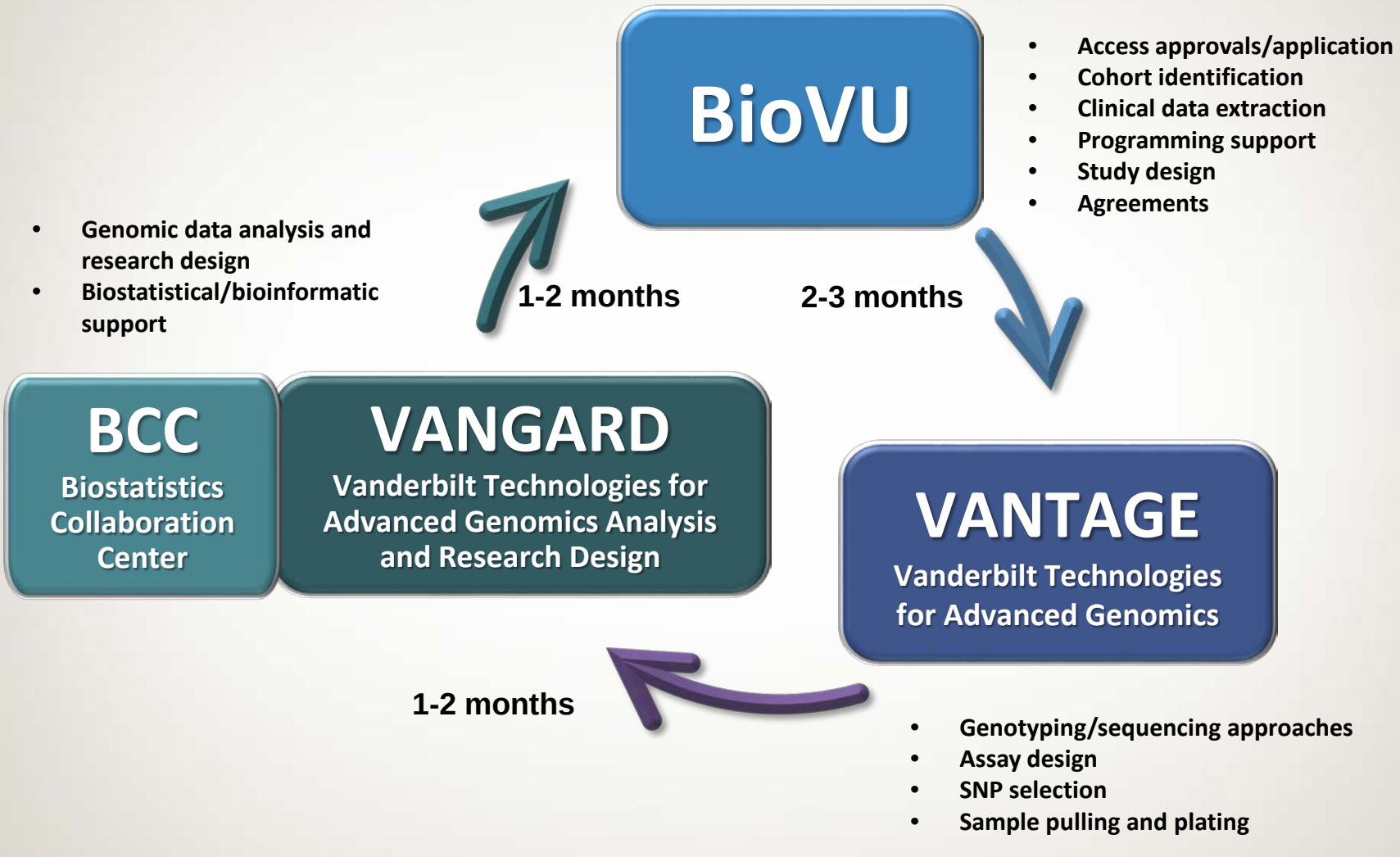
- DNA samples available: >180,000
- Plasma collection underway

Redeposited genotypes

- Subjects with GWAS data: >13,000
- Subjects with any genotyping: >70,000
- > 8,000,000,000 genotypes

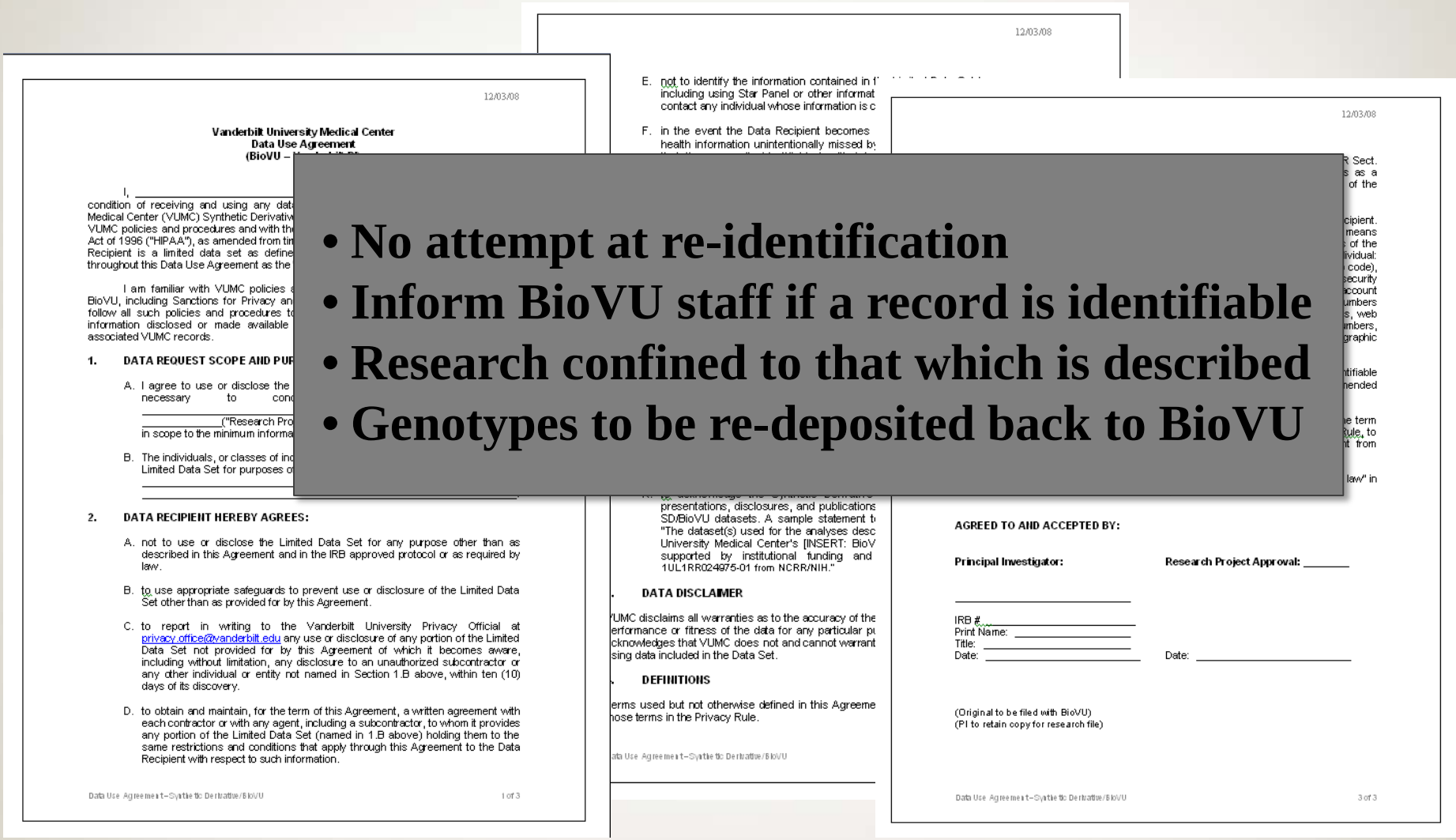
VANDERBILT UNIVERSITY
MEDICAL CENTER

BioVU Project Life Cycle



VANDERBILT UNIVERSITY
MEDICAL CENTER

Data Use Agreement



- No attempt at re-identification
- Inform BioVU staff if a record is identifiable
- Research confined to that which is described
- Genotypes to be re-deposited back to BioVU

VANDERBILT UNIVERSITY MEDICAL CENTER

Access to Existing Data

Home » BioVU & SD » BioVU Genotyping

BioVU Genotyping

# OF ADULTS in BioVU	# OF SAMPLES Genotyped	# OF CHILDREN in BioVU	# OF RECORDS in SD
167,342	61,523	23,123	2,609,958

One key goal of the BioVU resource is to enable exploration of the relationships among genetic variation, disease susceptibility, and variable drug responses.

As BioVU grows through investigator initiated genotyping, the resulting genotype data is continuously re-deposited back into the database for use by other investigators. Currently, you can search the following genetic data:

Targeted genotyping (for data obtained by targeted Taqman and Sequenom projects), **GWAS** (genome-wide genotyping), or high density genotyping **platform** (see below for an explanation and list of SNPs genotyped on the platforms).

Demographics Data

The table below shows the demographic data for currently requestable GWAS data.

Total
Female
Male
African-American
Asian
Caucasian
Hispanic
Other

Home » BioVU & SD » SNP Search Tool

Search for SNPs of Interest

The table below can be searched by rs number to provide the number of BioVU samples with genotyping data for that SNP broken down by genotyping platform. The number of unique samples with genotyping data for a particular SNP is reflected in the total.

Search:

Platform	Subject Count
ADME	8,467
HumanOmni5-Quad	2,018
HumanOmniExpressExome-8v1-A	761
Illumina 1M	1,489
Illumina 660W	3,376
Illumina Infinium HumanExome BeadChip	36,400
Illumina MetaboChip	14,099
Illumina OMNI-Quad	5,471
Illumina OmniExpress 1.0	128
Immunochip	3,365
Targeted	33,236
Total (unique subjects)	**65,822**

VANDERBILT UNIVERSITY
MEDICAL CENTER

Access to Existing Data

Home » BioVU & SD » Application Instructions » Existing Data Access

Existing Data Access

Components of a Data Access Application

Data Access Only: For investigators who wish to access the SD and existing genetic data.

1. **Download the Data Access Application Checklist.**

2. **Obtain IRB approval:**
   a. Complete IRB determination of non-human subject form #1122 on the IRB website and use the SD template language.
   b. Submit the form to IRB through DISCOVR-E, the IRB submission portal.
   c. You will receive an approval letter within 2 weeks - please retain this approval letter.

3. **As the Primary Investigator, compose the Research Proposal, Investigator Biosketch and Accept the Data Use Agreement.**

4. **Submit an Electronic Application.**

BioVU Research Proposal Checklists

- Statistical Analysis Plan Checklist
- Phenotype Plan Checklist

External Agreements

- External Data Use Agreement
- MTAShare- Materials Transfer Agreement Instructions
- MTAShare

Contact Us

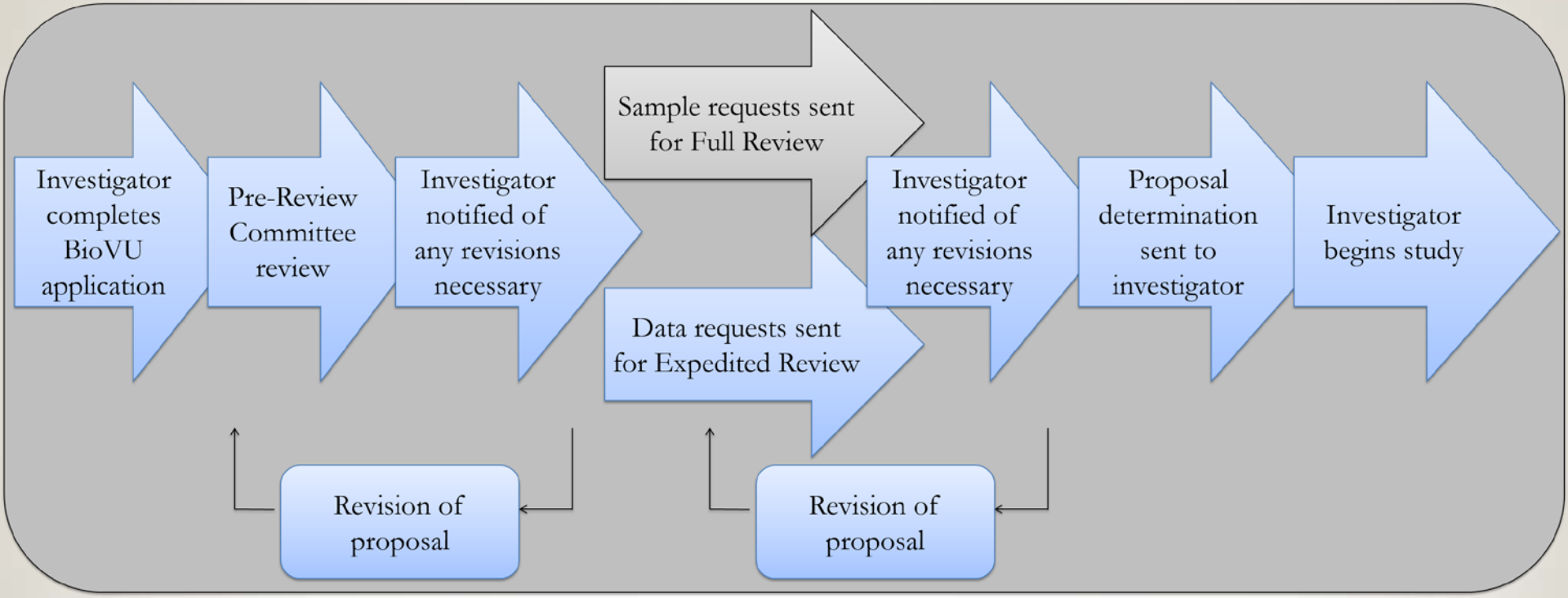
For more information about BioVU and the SD, contact the Team at biovu@vanderbilt.edu.

Make a Suggestion

Suggestion Form

VANDERBILT UNIVERSITY
MEDICAL CENTER

Access to Existing Data



VANDERBILT UNIVERSITY
MEDICAL CENTER

BioVU Sample Access

Sample Access

Componenets of a Sample Access Application

For investigators who will be requesting genotyping of BioVU DNA samples and access to the SD.

1. **Download the Sample Access Application Checklist.**

2. **Obtain IRB approval:**

   a. Complete IRB determination of non-human subject form #1122 on the IRB website and use the SD template language.

   b. Submit the form to IRB through DISCOVR-E, the IRB submission portal.

   c. You will receive an approval letter within 2 weeks - please retain this approval letter.

3. **As the Primary Investigator, compose the Research Proposal, Investigator Biosketch and Accept the Data Use Agreement.**

4. **Submit Electronic Application**

BioVU Research Proposal Checklists

Statistical Analysis Plan Checklist

Phenotype Plan Checklist

External Agreements

External Data Use Agreement

MTAShare- Materials Transfer Agreement Instructions

MTAShare

Contact Us

For more information about BioVU and the SD, contact the Team at biovu@vanderbilt.edu.

Make a Suggestion

Suggestion Form

VANDERBILT UNIVERSITY
MEDICAL CENTER

BioVU Sample Access

BioVU Proposal

Content of Research Plan (limit to 5 pages, not including references)

BioVU recommends the following format and page distribution. Organize *Items A-D* of the Research Plan to answer these questions: *What do you intend to do? Why is the work important? What has already been done? How are you going to do the work?*

A. Specific Aims

List the broad, long-term objectives and the goal of the specific research proposed, e.g., to test a stated hypothesis, create a novel design, solve a specific problem, challenge an existing paradigm or clinical practice, address a critical barrier to progress in the field, or develop new technology.

B. Background and Significance

Briefly sketch the background leading to the present application, critically evaluate existing knowledge, and specifically identify the gaps that the project is intended to fill. State concisely the importance and health relevance of the research described in this application by relating the specific aims to the broad, long-term objectives. If the aims of the application are achieved, state how scientific knowledge or clinical practice will be advanced. Describe the effect of these studies on the concepts, methods, technologies, treatments, services or preventative interventions that drive this field.

C. Preliminary Studies/Progress Report

Preliminary Studies. For new applications, use this section to provide an account of the principal investigator/program director's preliminary studies pertinent to this application. This information will also help to establish the experience and competence of the investigator to pursue the proposed project.

D. Research Design and Methods

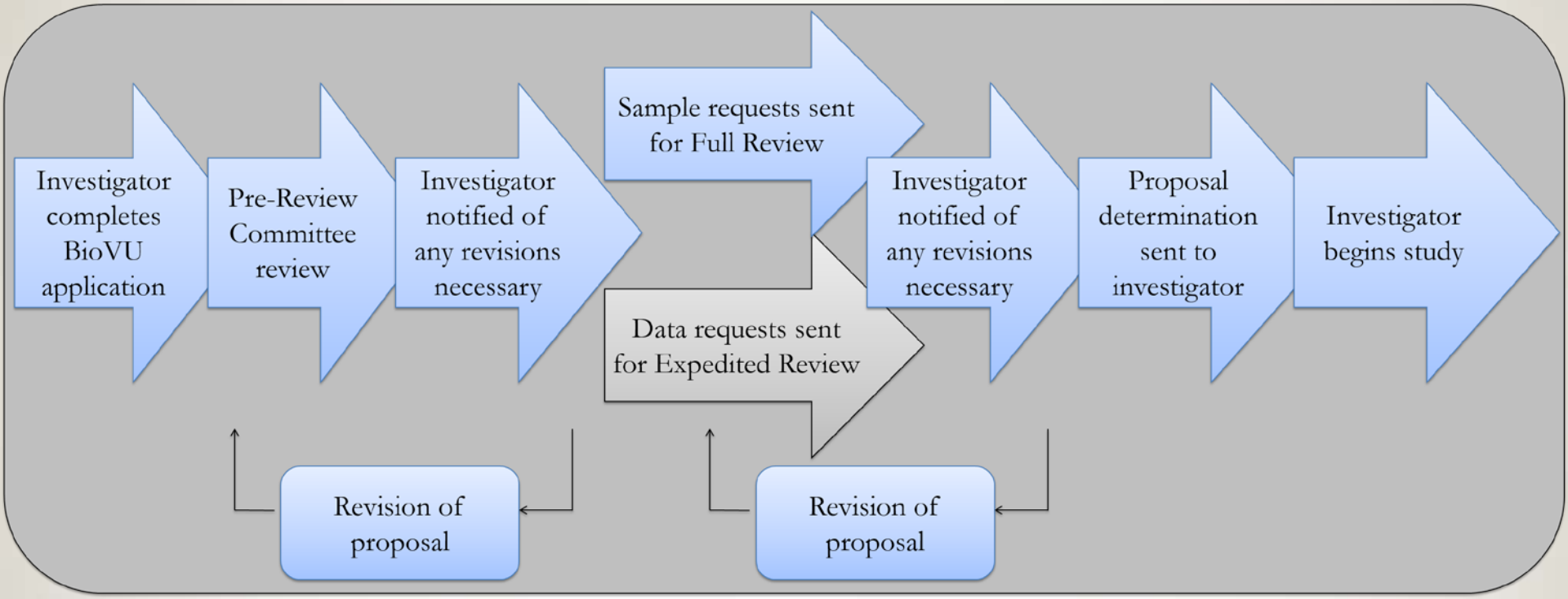
Describe the research design or conceptual framework, procedures, and analyses to be used to accomplish the specific aims of the project. Unless addressed separately in Section E, include how the data will be collected, analyzed, and interpreted as well as the data-sharing plan as appropriate. Fully describe the data set or phenotype of interest and methods for determining the data set that will be used. Describe any new methodology and its advantage over existing methodologies. Describe any novel concepts, approaches, tools, or technologies for the proposed studies. Discuss the potential difficulties and limitations of the proposed procedures and alternative approaches to achieve the aims. As part of this section, provide a tentative sequence or timetable for the project. Include the following:

1. SNP selection – list all SNPs to be genotyped and describe rationale for SNP selection.
2. Number of samples to be genotyped (include estimated number in BioVU).
3. DNA genotyping platform – describe anticipated platform (e.g. Sequenom, Taqman, etc.) and project size (# of chips, amount of DNA if applicable).
4. Resource requirements – include amount of DNA to be used.

E. Sample size justification and statistical analysis plan

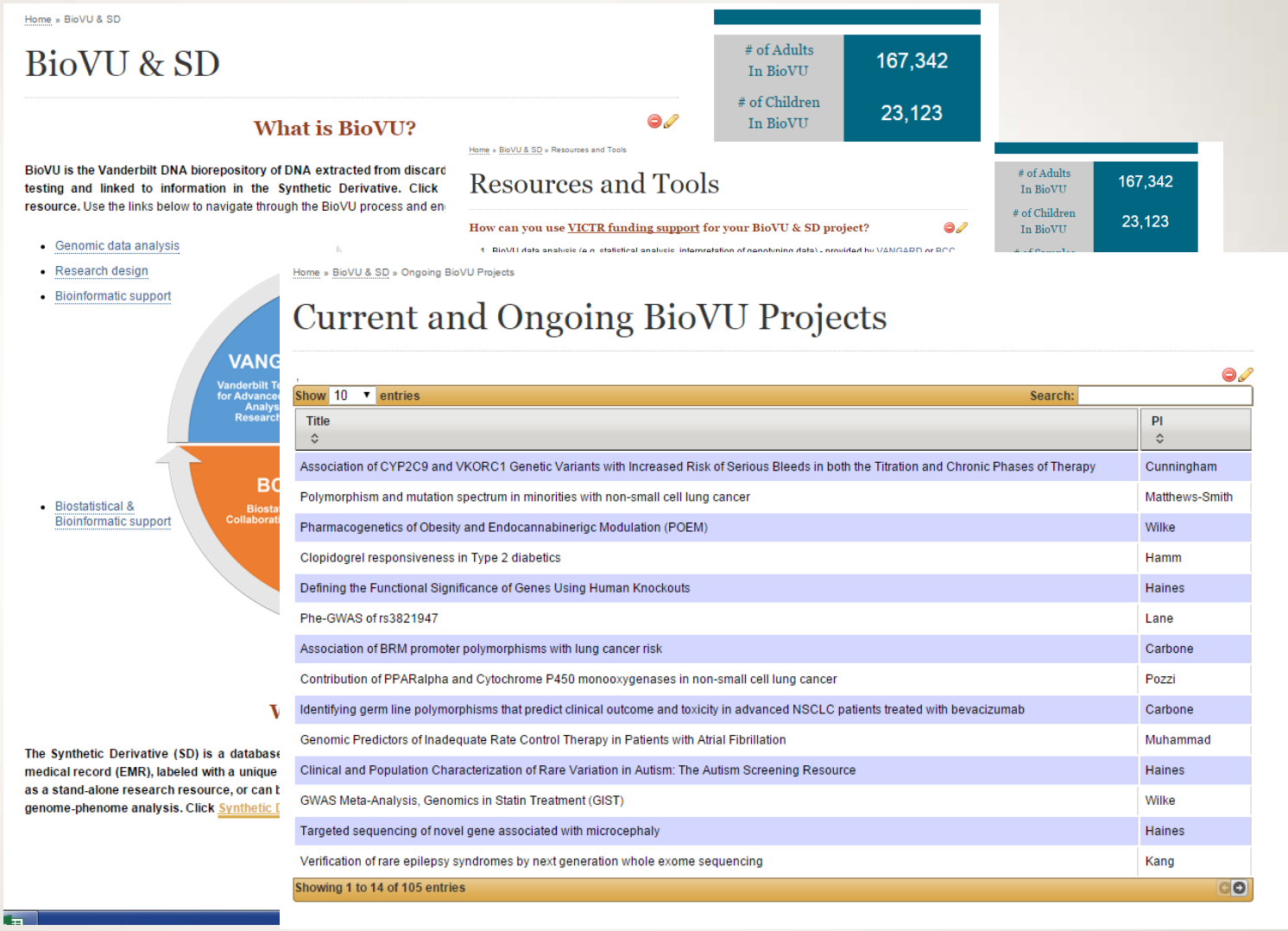
VANDERBILT UNIVERSITY
MEDICAL CENTER

BioVU Sample Access



VANDERBILT UNIVERSITY
MEDICAL CENTER

BioVU Website Update



VANDERBILT UNIVERSITY
MEDICAL CENTER

For ALL BioVU Studies…

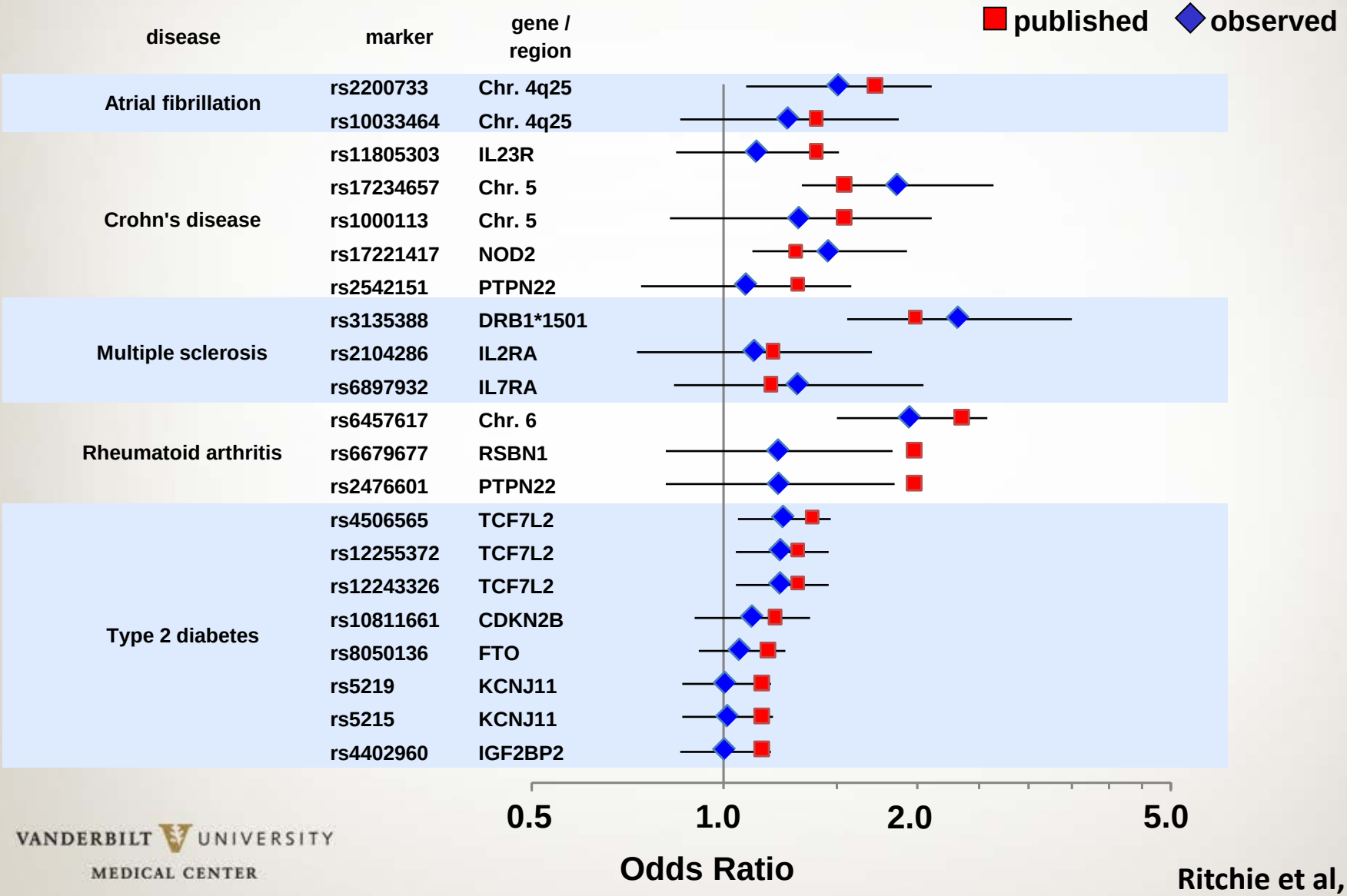
Resources:

1. **BioVU Project Management: BioVU@vanderbilt.edu**
2. **Programming services: IDASC CORE**
3. **Genomic technologies: VANTAGE CORE**
4. **Data analysis services: VANGARD CORE**

https://starbrite.vanderbilt.edu/biovu/

END

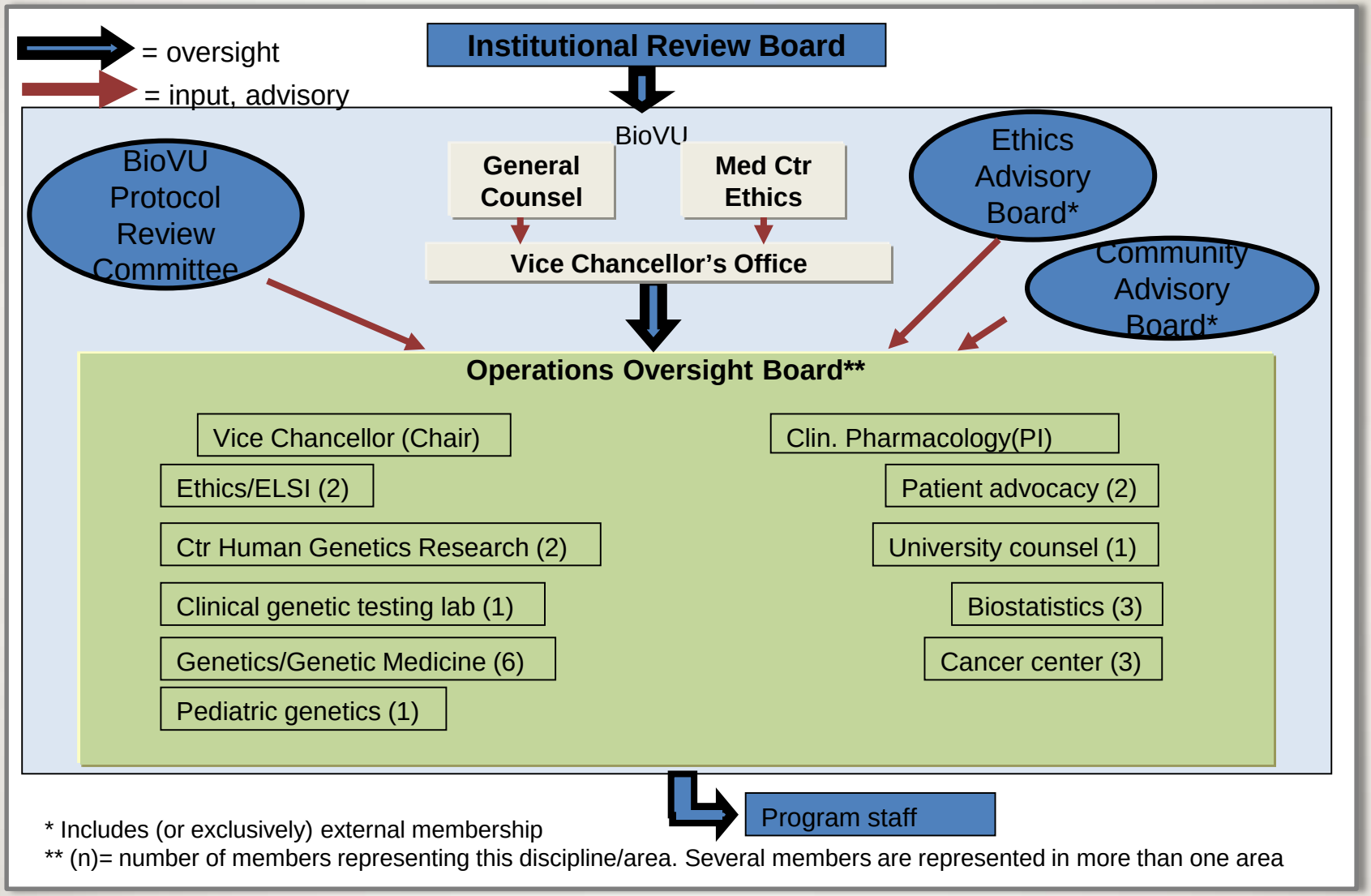
Validating EMR phenotype algorithms



VANDERBILT UNIVERSITY MEDICAL CENTER

Ritchie et al, 2010

BioVU Operations Oversight



* Includes (or exclusively) external membership

** (n)= number of members representing this discipline/area. Several members are represented in more than one area