



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!



BioVU Project Life Cycle



- Access approvals/application
- Cohort identification
- Clinical data extraction
- Programming support
- Study design
- Agreements



- Genomic data analysis and research design
- Biostatistical/bioinformatic support

2-3 months



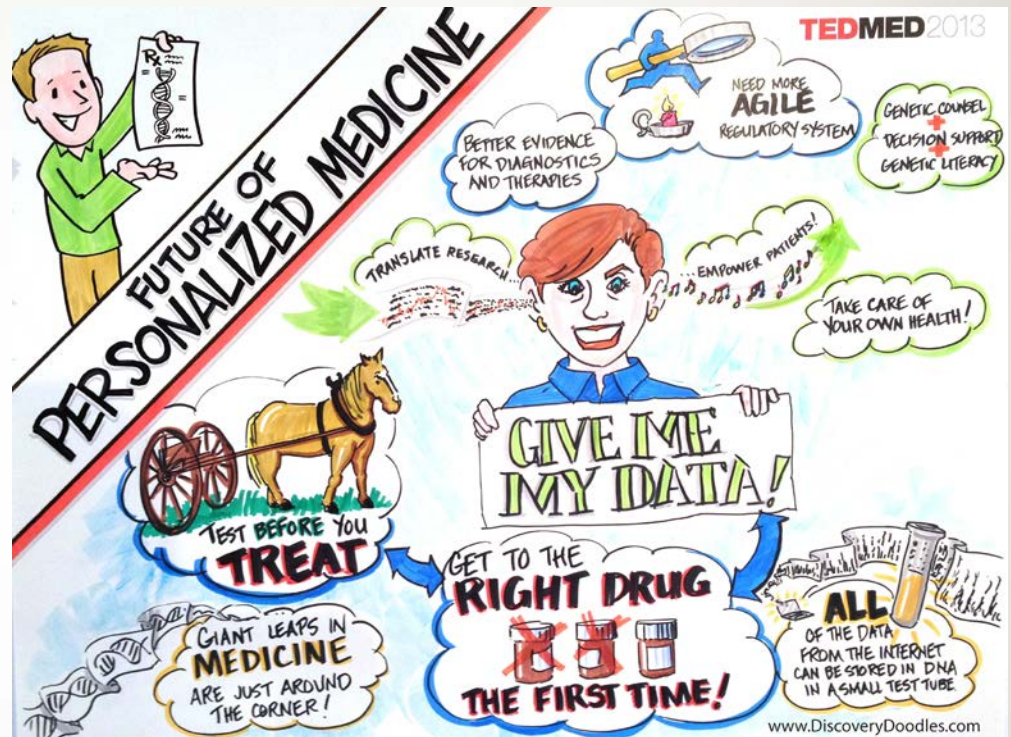
- Genotyping/sequencing approaches
- Assay design
- SNP selection
- Sample pulling and plating



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.



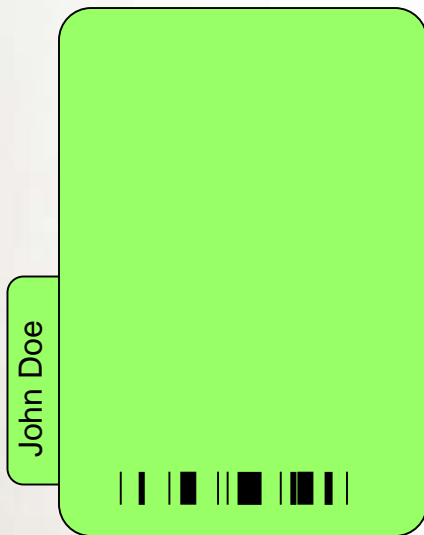


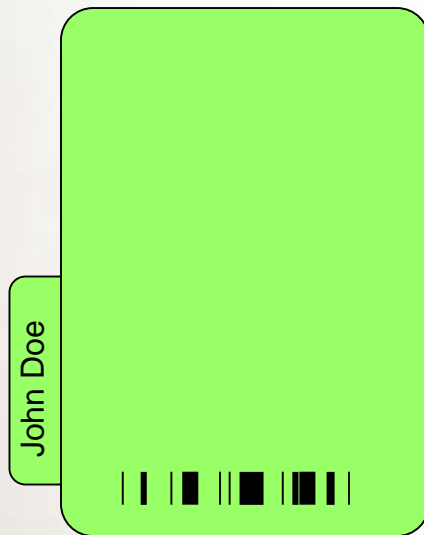
Resources for EMR-based research at VUMC

The Synthetic Derivative

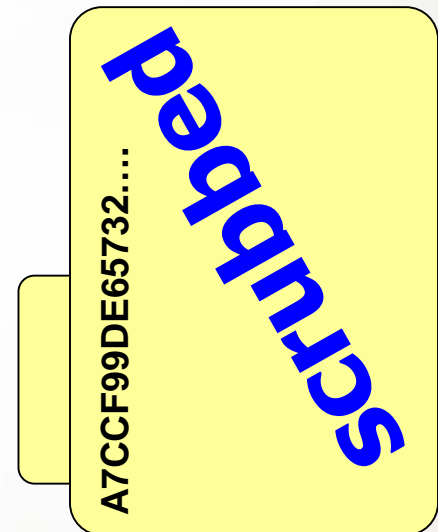
BioVU

Redeposited genotypes



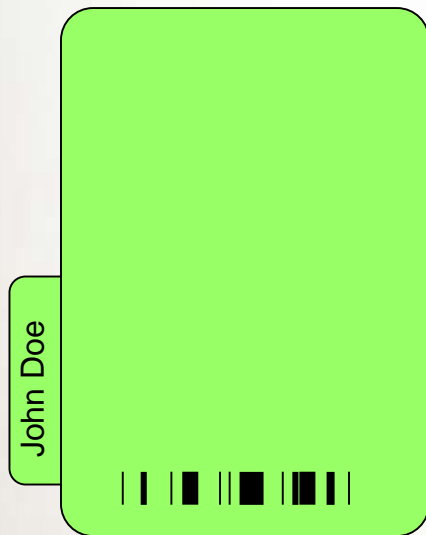


One way hash



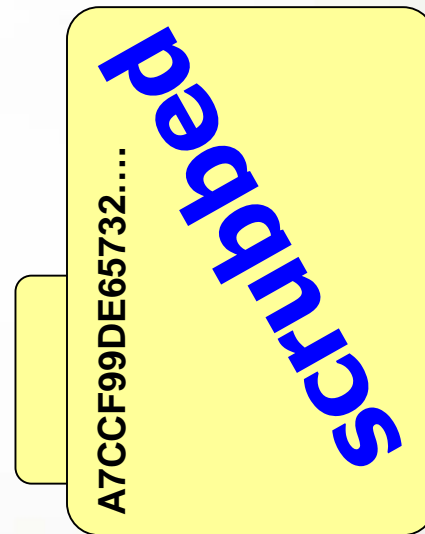
>2 million records

The Synthetic Derivative:
can be updated



→ eligible →

One way hash



>2 million records

The Synthetic Derivative:
can be updated



Extract
DNA
→





Resources for EMR-based research at VUMC

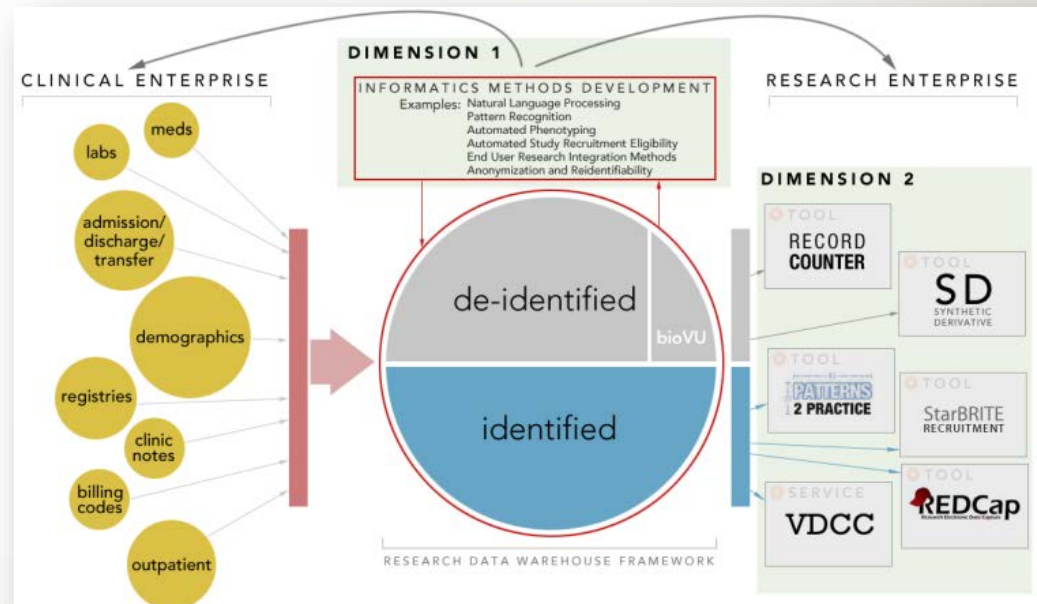
The Synthetic Derivative

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



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





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



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

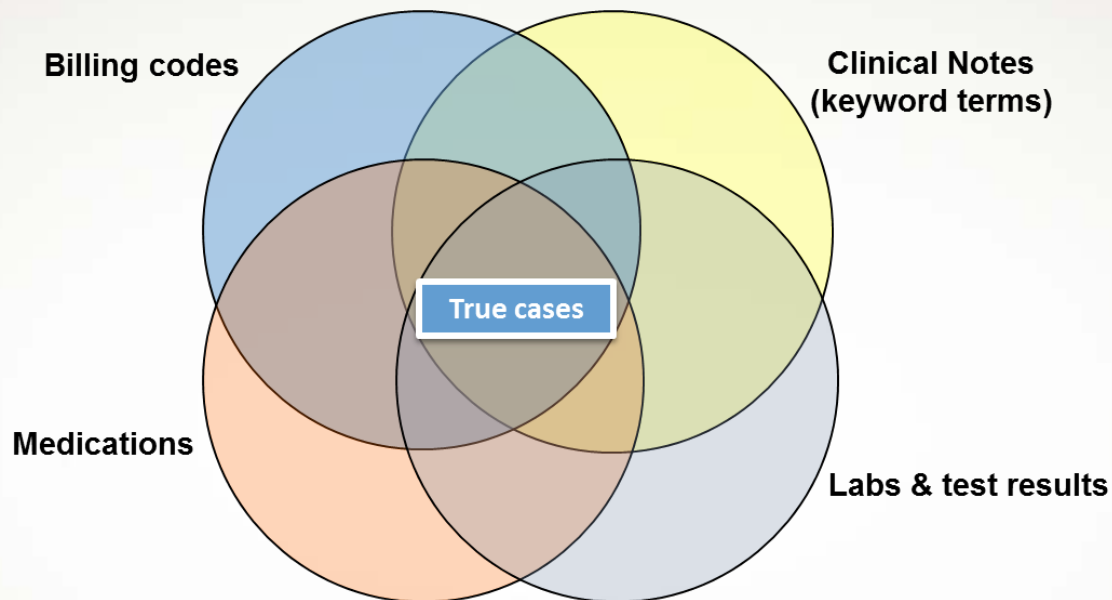
- Our algorithm shifts the dates within a record by a time period (*up to 364 days backwards*) that is consistent within each record, but differs *across* records

Restricted access & continuous oversight

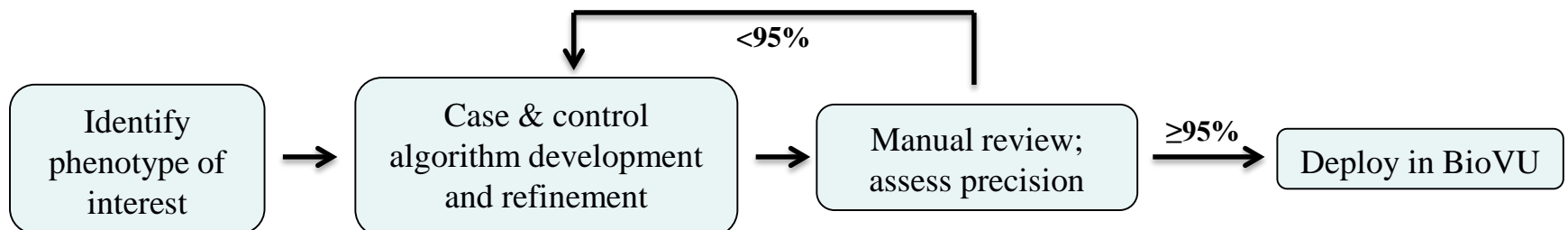
- Access restricted to VU; not a public resource
- IRB approval for study (non-human)
- Data Use Agreement
- Audit logs of all searches and data exports



Phenotyping Approach




Algorithm Development






Feasibility Counts Record Counter Tool

[UBER](#) [MY RESEARCH](#) [GOV. DASHBOARD](#) [PUB. COMPLIANCE](#) [MY PROFILE](#) [QUICK LINKS](#) [LOGOUT](#) [CONTACT US](#)

VANDERBILT  UNIVERSITY


StarBRITE

Search StarBRITE 

[Planning & Implementation](#) [Recruitment](#) [Research on Practice and Policy](#) [Funding](#) [Data Management](#) [Education](#) [BioVU & SD](#) [Basic Science](#)

[Home](#) » [BioVU & SD](#) » Record Counter

Record Counter

The Synthetic Derivative Record Counter (RecordCounter) provides exploratory data figures to members of the VU  research community for research planning purposes and feasibility assessment. The Record Counter allows the user to input basic medical data, such as ICD 9 codes or text keywords, e.g., lung cancer, as well as demographic information, and then search the Synthetic Derivative Database to determine the approximate number of records that meet those criteria.

RecordCounter 3.0 has just been launched. The RC 3.0 leverages the power of an IBM Netezza data warehouse appliance to provide faster, near-immediate counts as the user builds their search criteria. In addition, we have developed new features to improve the search and presentation capabilities.

- The user will now be able to get counts for each specific criteria element as denoted to the right hand side of the search box, summary counts for combined criteria (this OR that) indicated at the bottom of the group box, and a final Set Total count.
- Medications can now be selected by checking the box next to the medication name and selecting Add Criteria. This allows you to select multiple medications containing the search term at once.
- Detailed results of the query can be viewed and exported after saving the set.
- Multiple mentions of ICD-9 codes can be specified.

Check out our [PowerPoint tutorial](#) for more information about the new updates to the RecordCounter 3.0.

Firefox is the preferred browser for using Record Counter.
[Tips for using the Record Counter](#)

★ Open Record Counter

# of Adults In BioVU	167,342
# of Children In BioVU	23,123
# of Samples With GWAS Data In BioVU	13,031

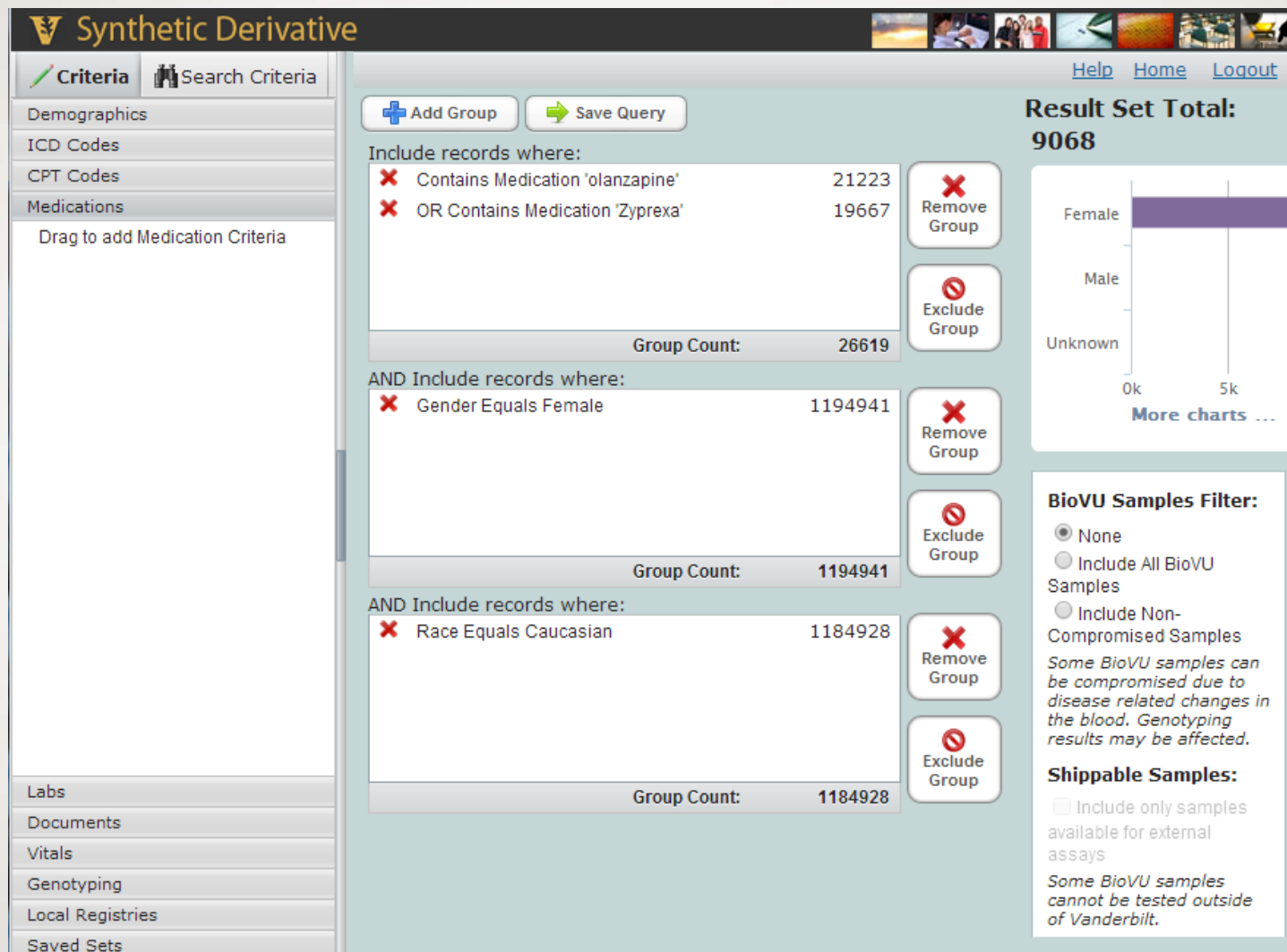


Contact Us

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

Make a Suggestion

[Suggestion Form](#)





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

- Non-Human Subjects Determination Letter
- Signed Data Use Agreement
- Complete the Online Application

Contact Us

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

Make a Suggestion

[Suggestion Form](#)

For Primary Investigators:

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



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



BioVU Sample Acceptance

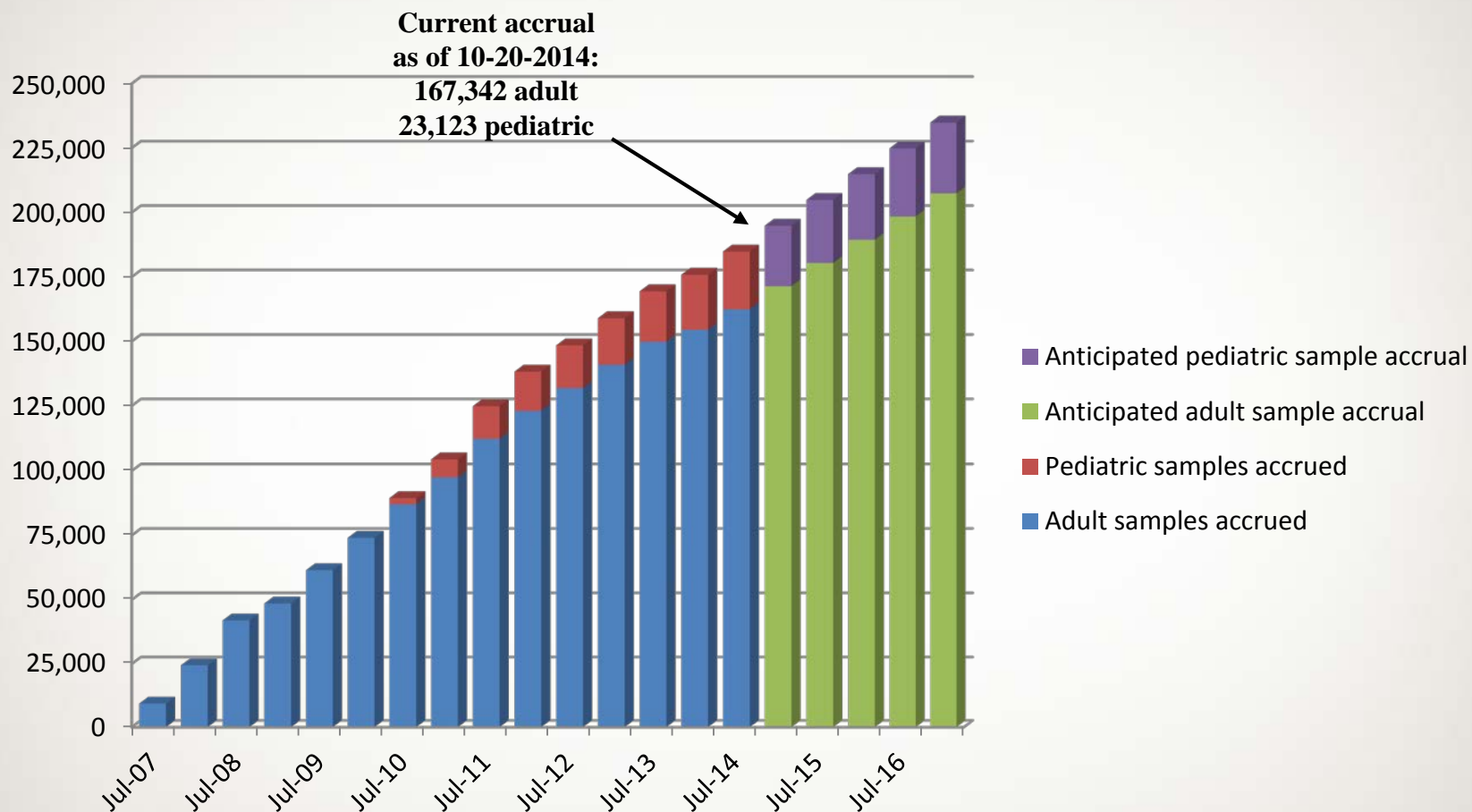


Accepted samples must:

- 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



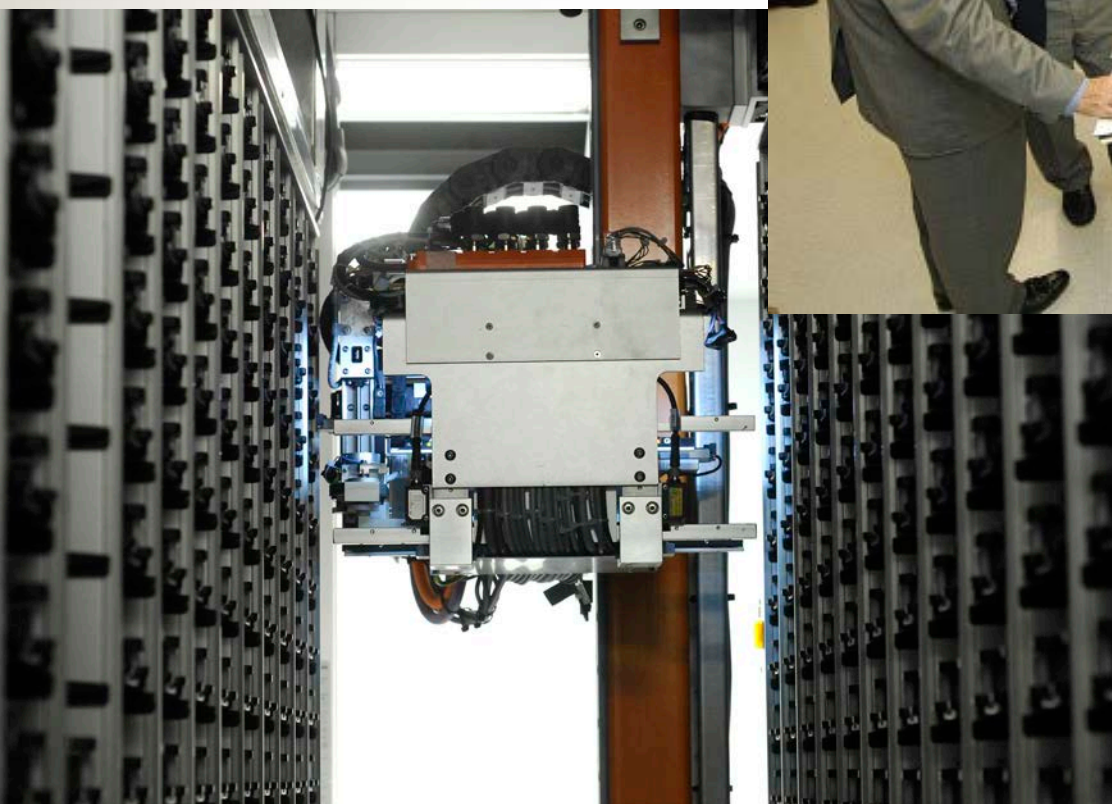
BioVU Sample Accrual: 190,336





VANDERBILT UNIVERSITY
MEDICAL CENTER

BioVU Storage



RTS SmaRTStore



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.

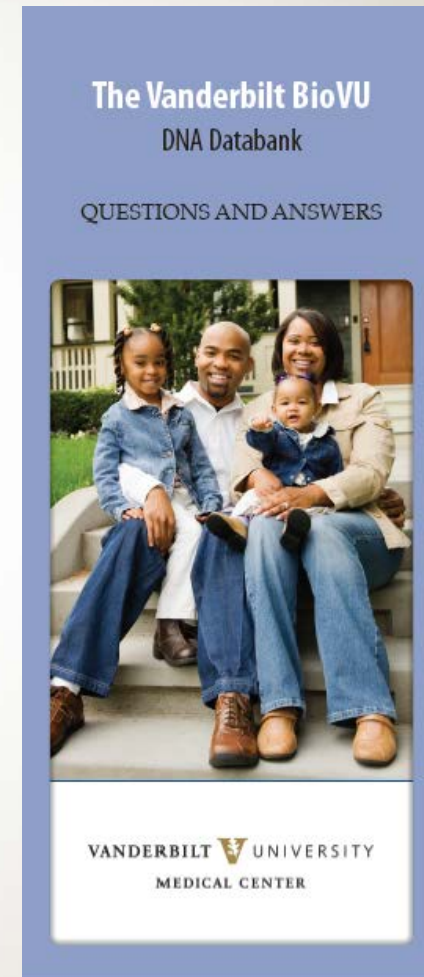
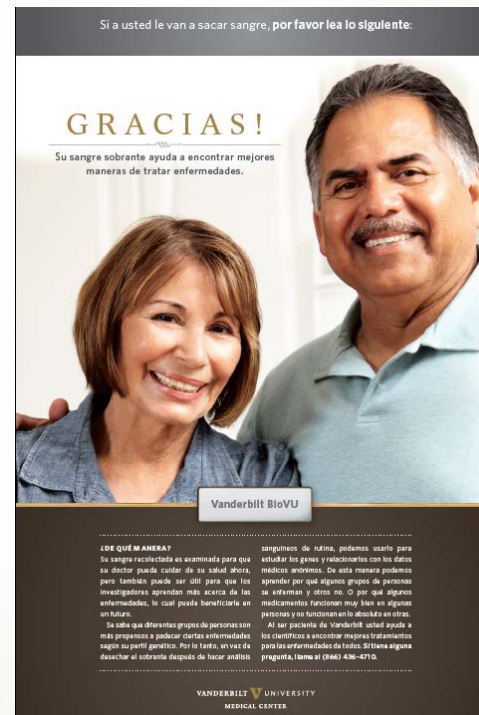
Patient/Legal Representative _____ Date _____

(Relationship to Patient) _____



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





Disease Cohorts

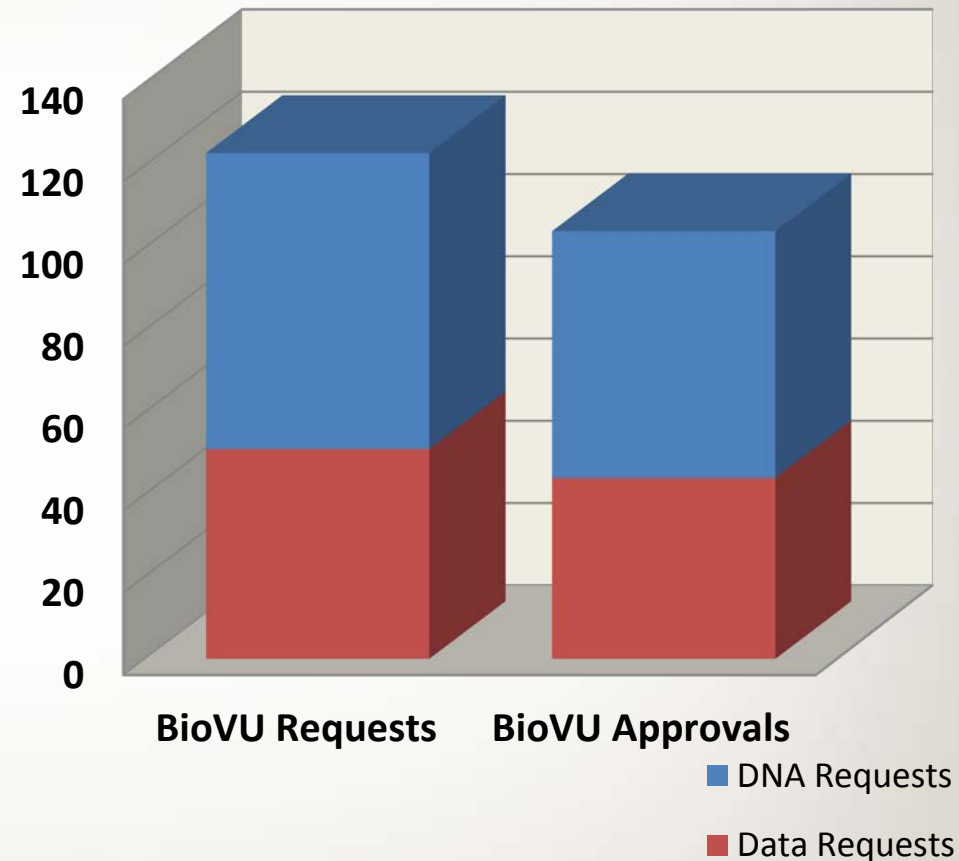
Rare Diseases

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



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**



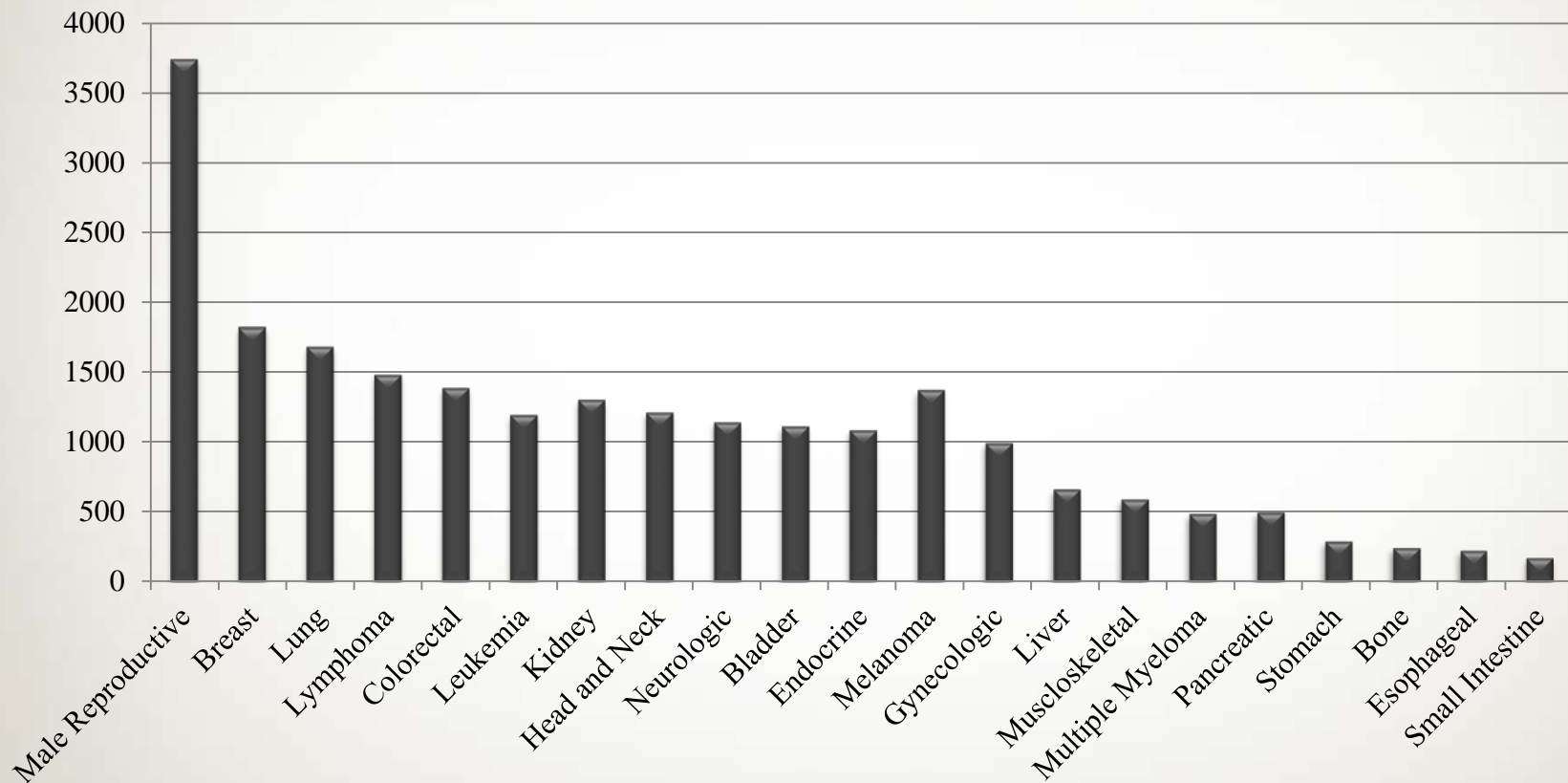
*coming Nov 1st, changes
to Expedited Review



Disease Cohorts

Oncology

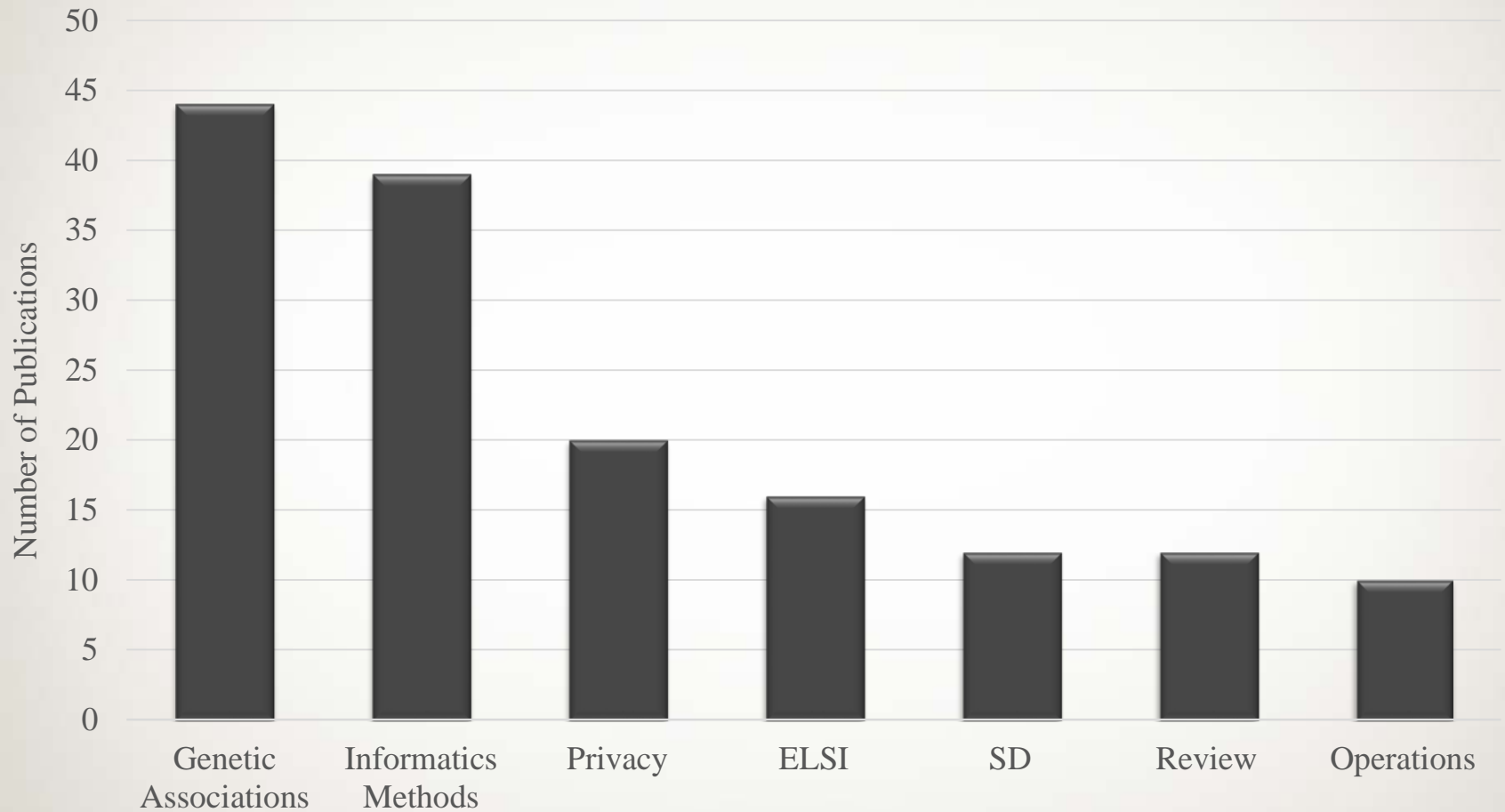
Most Common Cancers in BioVU*



*as defined by Tumor Registry, September 2014

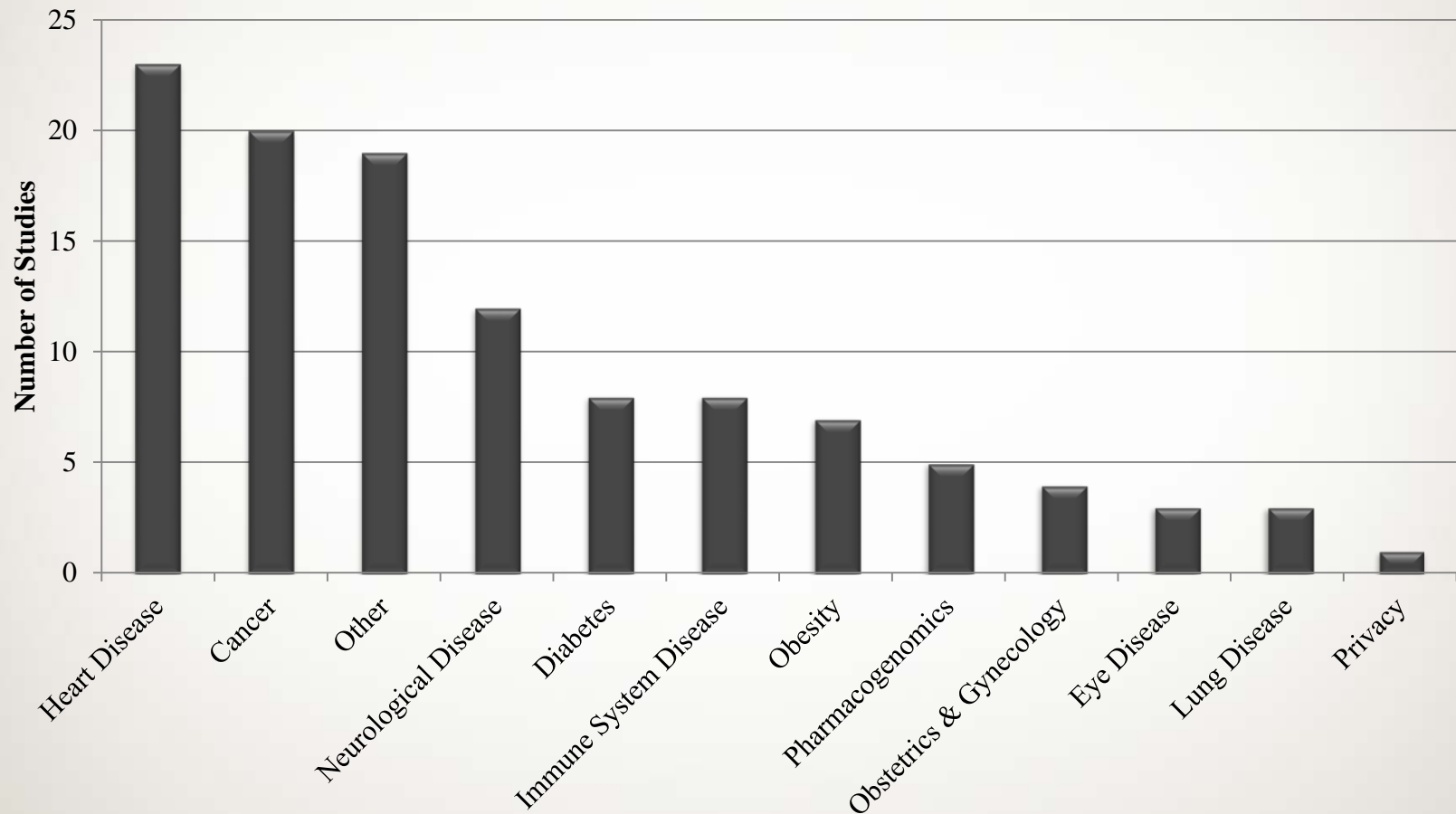


BioVU and SD Publication Topics





Breadth of BioVU Studies





Resources for EMR-based research at VUMC

The Synthetic Derivative

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

BioVU

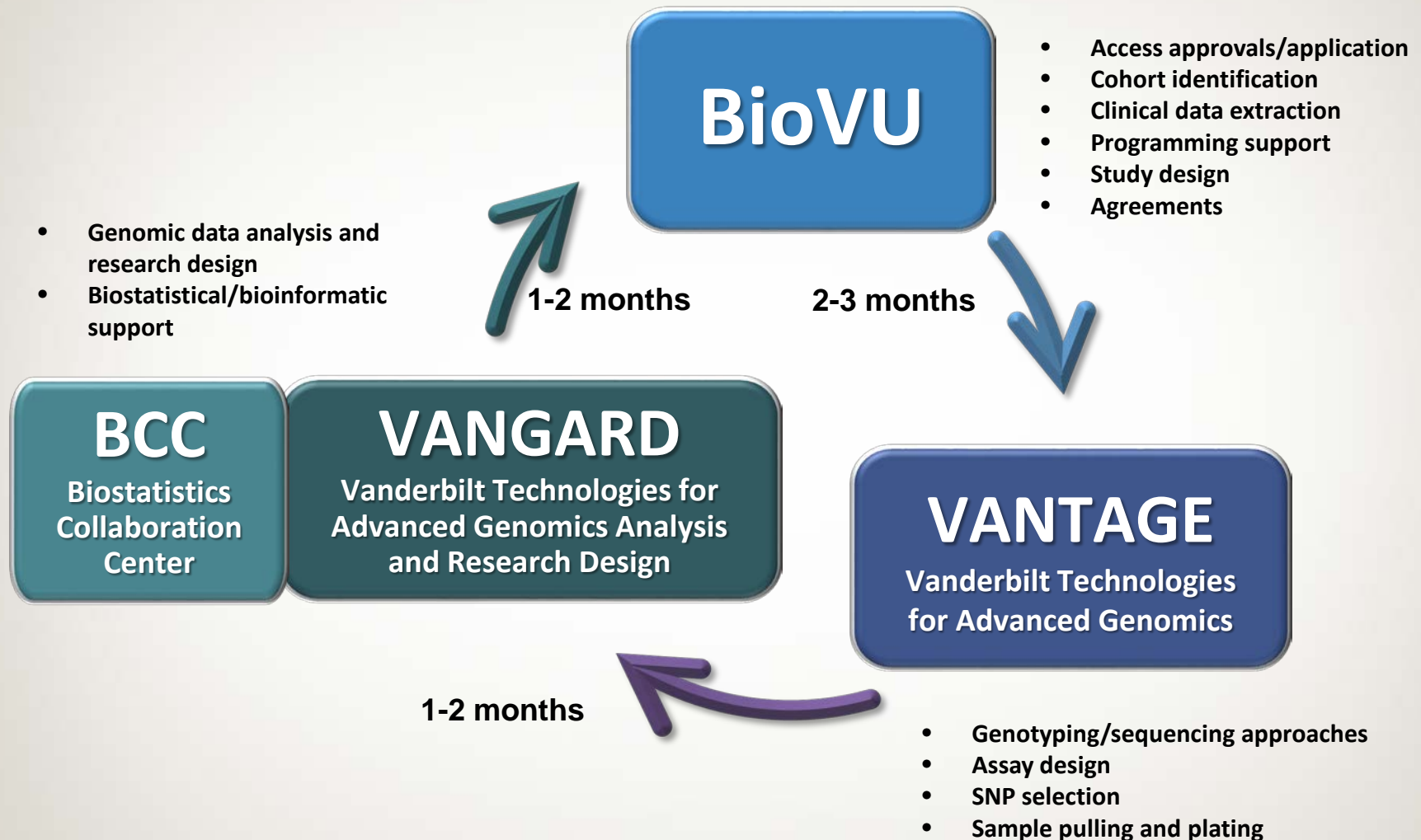
- DNA samples available: >180,000
- Plasma collection underway

Redeposited genotypes

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



BioVU Project Life Cycle





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
Data Use Agreement
(BioVU – Synthetic Derivative)**

12/03/08

12/03/08

12/03/08

I, _____
condition of receiving and using any data from the
Medical Center (VUMC) Synthetic Derivative (SD) datasets.
VUMC policies and procedures and with the
Act of 1996 ("HIPAA"), as amended from time to time.
Recipient is a limited data set as defined in the Privacy
throughout this Data Use Agreement as the "Limited Data Set."

I am familiar with VUMC policies and procedures for
BioVU, including Sanctions for Privacy and Security.
I will follow all such policies and procedures to protect
information disclosed or made available to me from
associated VUMC records.

1. DATA REQUEST SCOPE AND PURPOSE

A. I agree to use or disclose the Limited Data Set for
necessary to conduct the research project described in
this agreement.

("Research Project")
in scope to the minimum information necessary to
conduct the research project.

B. The individuals, or classes of individuals, included in the
Limited Data Set for purposes of this agreement are:

2. DATA RECIPIENT HEREBY AGREES:

A. not to use or disclose the Limited Data Set for any purpose other than as
described in this Agreement and in the IRB approved protocol or as required by
law.

B. to use appropriate safeguards to prevent use or disclosure of the Limited Data
Set other than as provided for by this Agreement.

C. to report in writing to the Vanderbilt University Privacy Official at
privacy.office@vanderbilt.edu any use or disclosure of any portion of the Limited
Data Set not provided for by this Agreement of which it becomes aware,
including without limitation, any disclosure to an unauthorized subcontractor or
any other individual or entity not named in Section 1.B above, within ten (10)
days of its discovery.

D. to obtain and maintain, for the term of this Agreement, a written agreement with
each contractor or with any agent, including a subcontractor, to whom it provides
any portion of the Limited Data Set (named in 1.B above) holding them to the
same restrictions and conditions that apply through this Agreement to the Data
Recipient with respect to such information.

E. not to identify the information contained in the Limited Data Set
including using Star Panel or other information that could
contact any individual whose information is contained in the Limited Data Set.

F. in the event the Data Recipient becomes aware of health information
unintentionally missed by the Data Recipient, to promptly notify the
Vanderbilt University Medical Center.

presentations, disclosures, and publications derived from the
SD/BioVU datasets. A sample statement to the IRB might be:
"The dataset(s) used for the analyses described in this agreement
University Medical Center's (INSERT: BioVU Synthetic Derivative)
supported by institutional funding and NCI/NCRR/NIH."

DATA DISCLAIMER

VUMC disclaims all warranties as to the accuracy or fitness of the data for any particular purpose.
The Data Recipient acknowledges that VUMC does not and cannot warrant the accuracy or fitness of
sing data included in the Data Set.

DEFINITIONS

Terms used but not otherwise defined in this Agreement shall be defined in the Privacy Rule.

Data Use Agreement – Synthetic Derivative/BioVU

AGREED TO AND ACCEPTED BY:

Principal Investigator: _____

Research Project Approval: _____

IRB #: _____
Print Name: _____
Title: _____
Date: _____

Date: _____

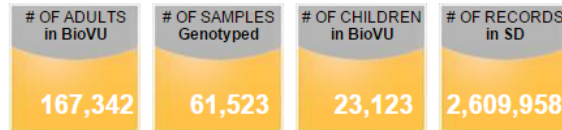
(Original to be filed with BioVU)
(PI to retain copy for research file)



Access to Existing Data

[Home](#) » [BioVU & SD](#) » BioVU Genotyping

BioVU Genotyping



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).

[Home](#) » [BioVU & SD](#) » SNP Search Tool

Demographics Data

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

Total
Female
Male
African-American
Asian
Caucasian
Hispanic
Other

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



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 [DISCOVER-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

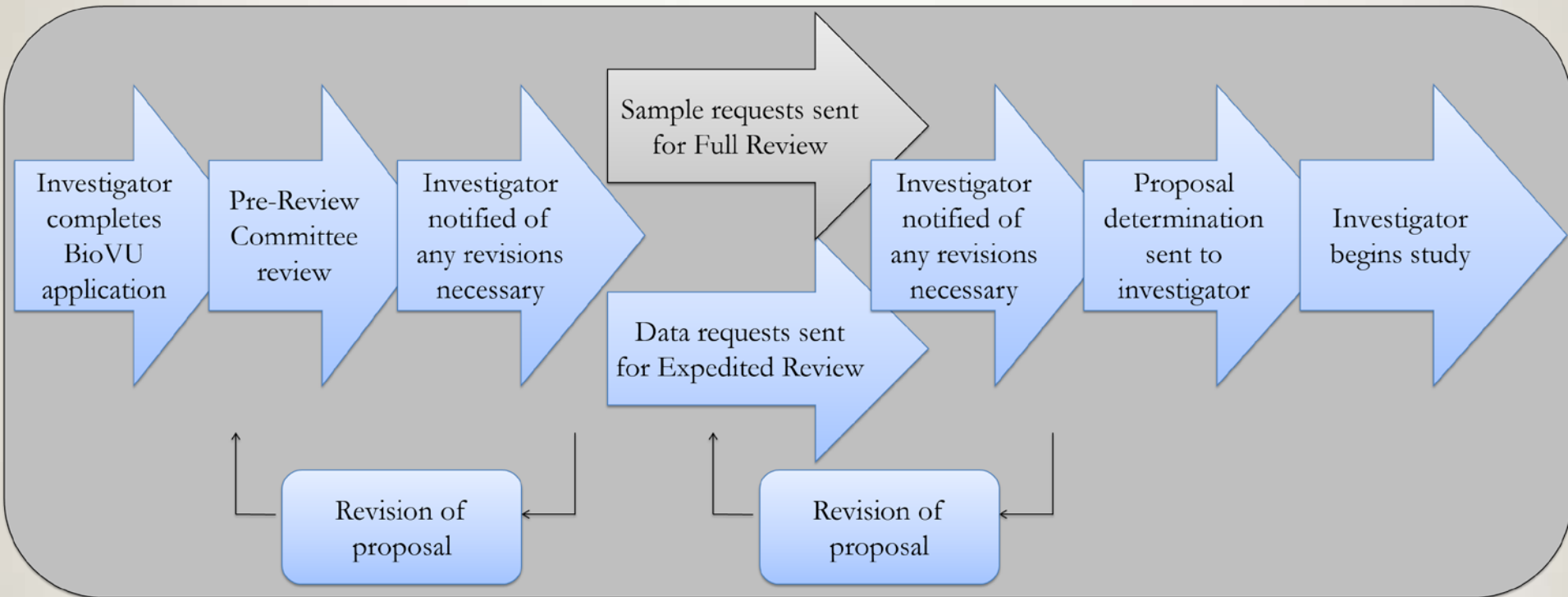
For more information about BioVU and the SD, contact the Team at biovu@vanderbilt.edu.

Make a Suggestion

[Suggestion Form](#)



Access to Existing Data





BioVU Sample Access

[Home](#) » [BioVU & SD](#) » [Application Instructions](#) » [Sample Access](#)

Sample Access

Components 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 [DISCOVER-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](#)



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

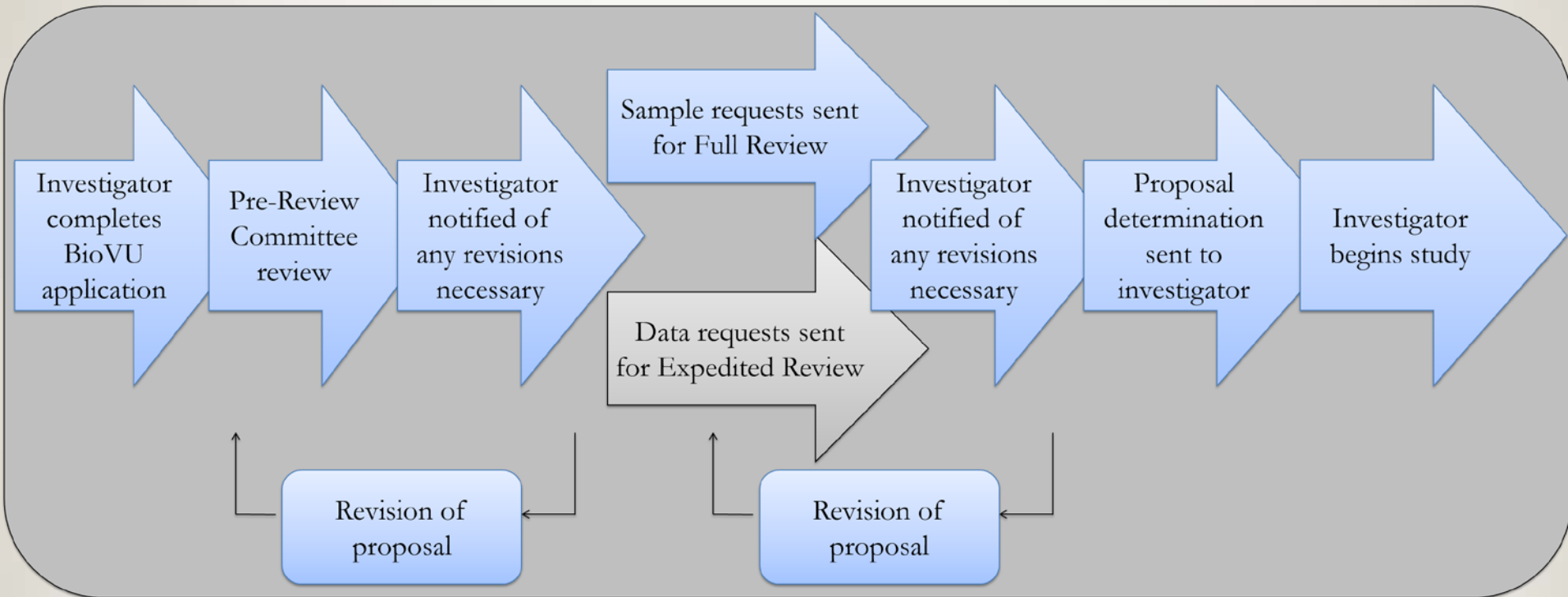
Describe the research design or conceptual framework, procedures, and analyses to be used to accomplish the specific aims of the project. Unless addressed separately in Section E, include how the data will be collected, analyzed, and interpreted as well as the data-sharing plan as appropriate. Fully 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



BioVU Sample Access





BioVU Website Update

Home » BioVU & SD

BioVU & SD

What is BioVU?

BioVU is the Vanderbilt DNA biorepository of DNA extracted from discard testing and linked to information in the Synthetic Derivative. Click resource. Use the links below to navigate through the BioVU process and en

- Genomic data analysis
- Research design
- Bioinformatic support
- Biostatistical & Bioinformatic support

Home » BioVU & SD » Resources and Tools

Resources and Tools

How can you use **VICTR funding support** for your BioVU & SD project?

1 BioVU data analysis (a n statistical analysis, interpretation of nonhuman data) - provided by VANGARD or BCC

Home » BioVU & SD » Ongoing BioVU Projects

Current and Ongoing BioVU Projects

Show 10 entries Search:

Title	PI
Association of CYP2C9 and VKORC1 Genetic Variants with Increased Risk of Serious Bleeds in both the Titration and Chronic Phases of Therapy	Cunningham
Polymorphism and mutation spectrum in minorities with non-small cell lung cancer	Matthews-Smith
Pharmacogenetics of Obesity and Endocannabinergic Modulation (POEM)	Wilke
Clopidogrel responsiveness in Type 2 diabetics	Hamm
Defining the Functional Significance of Genes Using Human Knockouts	Haines
Phe-GWAS of rs3821947	Lane
Association of BRM promoter polymorphisms with lung cancer risk	Carbone
Contribution of PPARalpha and Cytochrome P450 monooxygenases in non-small cell lung cancer	Pozzi
Identifying germ line polymorphisms that predict clinical outcome and toxicity in advanced NSCLC patients treated with bevacizumab	Carbone
Genomic Predictors of Inadequate Rate Control Therapy in Patients with Atrial Fibrillation	Muhammad
Clinical and Population Characterization of Rare Variation in Autism: The Autism Screening Resource	Haines
GWAS Meta-Analysis, Genomics in Statin Treatment (GIST)	Wilke
Targeted sequencing of novel gene associated with microcephaly	Haines
Verification of rare epilepsy syndromes by next generation whole exome sequencing	Kang

Showing 1 to 14 of 105 entries



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

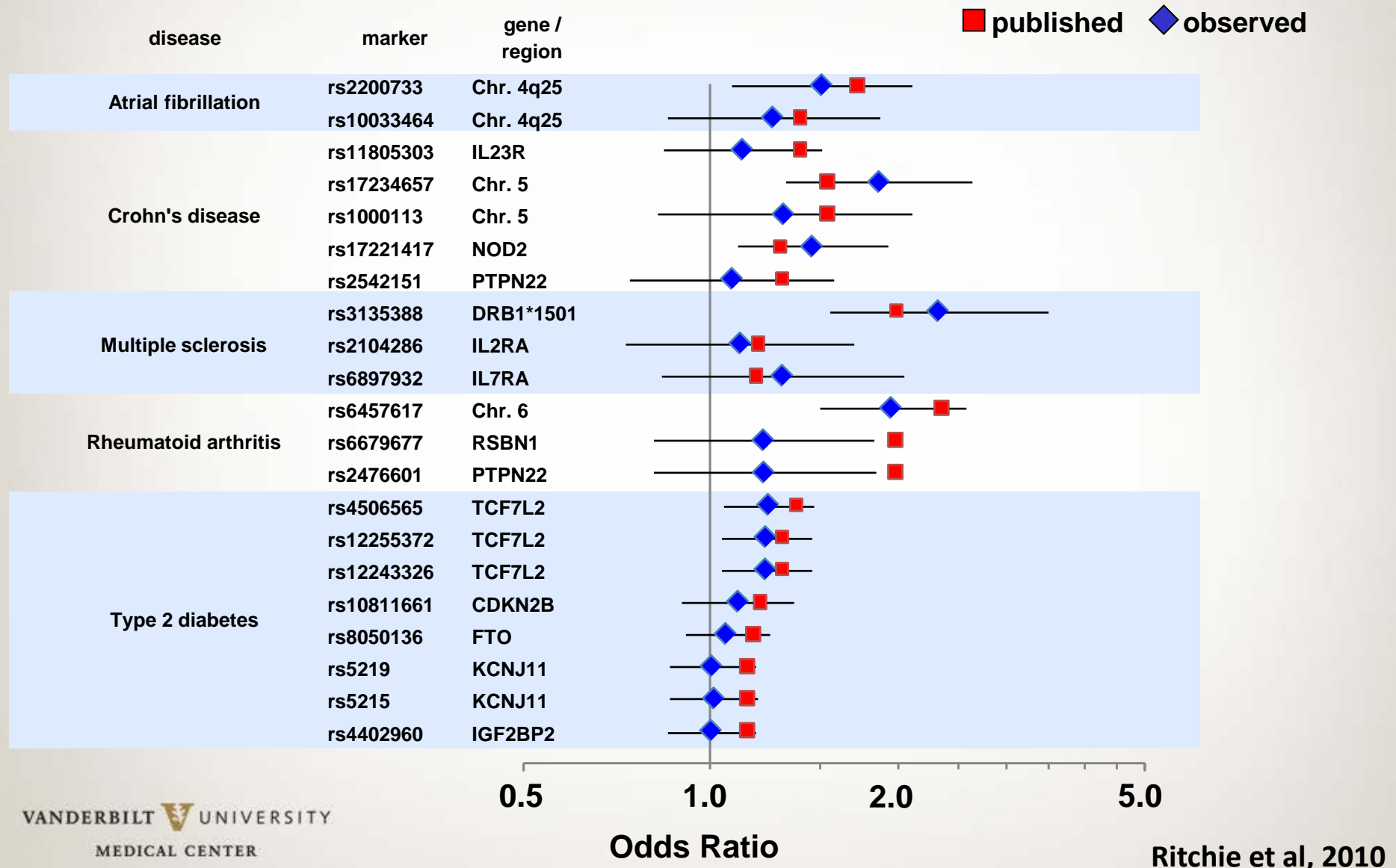
<https://starbrite.vanderbilt.edu/biovu/>



VANDERBILT UNIVERSITY
MEDICAL CENTER

END

Validating EMR phenotype algorithms



BioVU Operations Oversight

