Performing the best science: Pathways to validity and reliability and effective data managament

Christopher Ritchlin MD, MPH.
Professor of Medicine
Chief, Allergy, Immunology & Rheumatology Division
Center for Musculoskeletal Research

MEDICINE *of* THE HIGHEST ORDER

UNIVERSITY *of* ROCHESTER MEDICAL CENTER

Evidence Based Science

1. Begley CG et al. Institutions must do their part for reproducibility. Nature 2015; 525:25

2. McCoun R et al Hide results to find the truth. Nature 2015; 526:187

3. Blumberg RS et al. Unravelling the autoimmune translational research process layer by layer. Nature Med 2012; 18(1):35

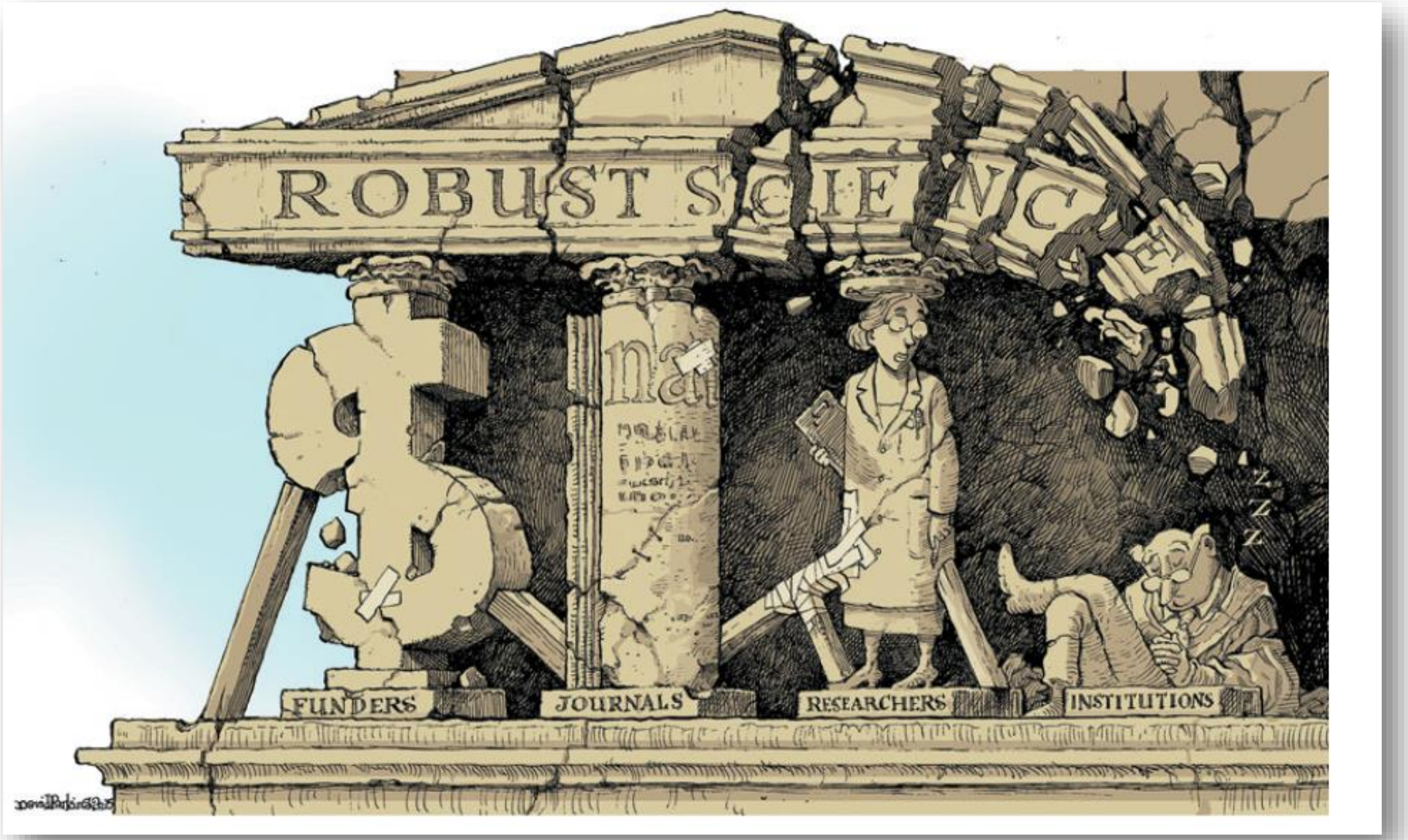
4. Kilkenny C. The ARRIVE Guidelines. PLOS Biology 2010

Overview

1. Reproducibility & Scientific Rigor
2. Data Collection and Storage
3. Data Analysis

Reproducibility and scientific rigor

Rigor and Reproducibility



<25% high profile science is reproducible

Journal Checklists
Biostatistician reviewers
NIH initiatives
No limit on methods
Stress quality over quantity

Begley G. Nature Sept 2015

Long Term Outcomes

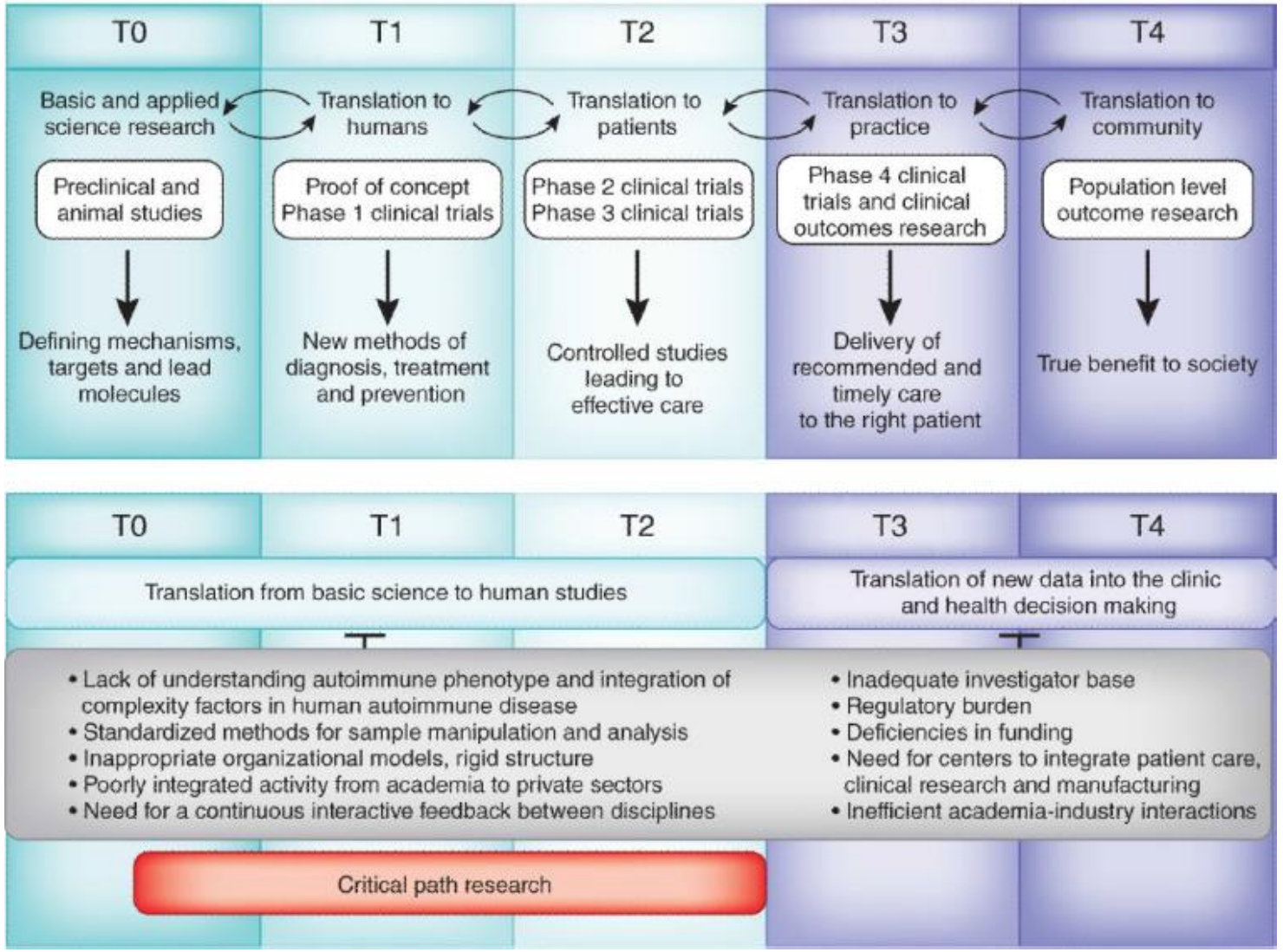

BIG DATA
Volume
Useful
Metadata
AMP
Structured data
Structured
Petabyte analysis
Unstructured content
People driven
Decision making
Data
Zettabyte
Framework
Smart content
database
Concept extraction
Text analytics
Semantic

Good Institutional Practice

- **Routine discussion of research methods**
- **Good reporting systems**
- **Training & standards**
- **Records & quality management**
- **Appropriate incentive & evaluation systems**

Begley GC. Nature Sept 2015

Operational challenges for translational medicine

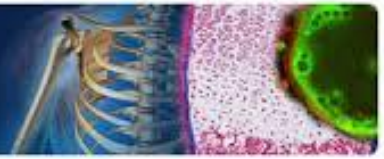


Blumberg RS. Nature Medicine 2012. 18(1):35

Establish Good Research Practices

Quality Research in Preclinical and Translational Research Core

Data Management, Bioinformatics and Analytics Core



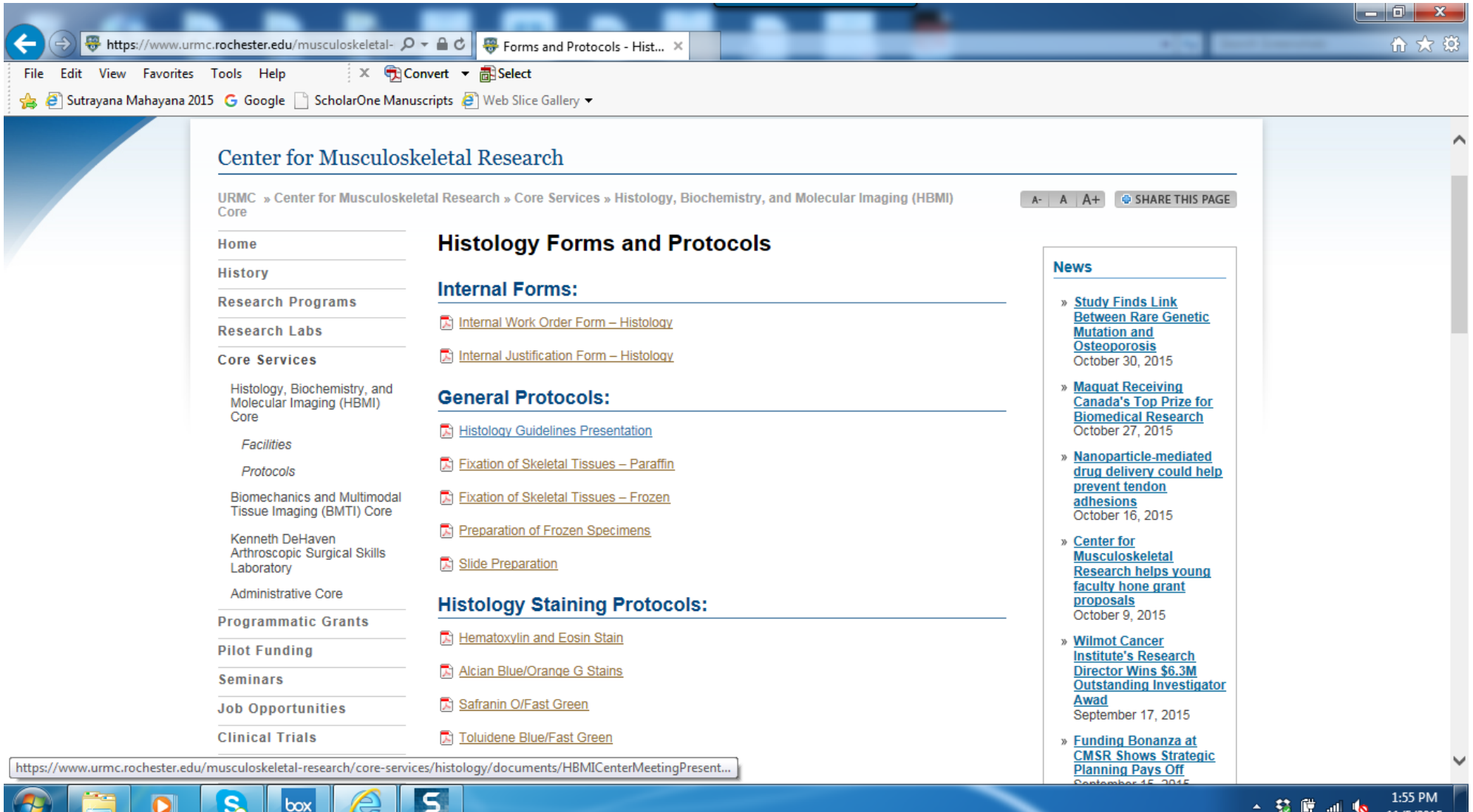
Standard Operating Procedures

Table 2. SOP Format and Content Description

Title: A specific name assigned to the document
Purpose: What the document will be used for
Scope: Describes boundaries of what procedure will / will not deliver
Responsibilities: Defines the roles and associated duties
Safety Precautions: Lists any specific safety precautions for procedure
Materials: Lists equipment, supplies and reagents with specifications
Specimens: Lists specimens, collection and storage requirements
Quality Control: Activities use to fulfill and verify quality requirements
Procedure: Sequence of steps defining what is to be done
Limitations: Items restricting execution when all steps are completed
Calculations: A process for transforming inputs into results
Interpretation: Information relating to the evaluation of results
Result reporting: Details how to report results
Training: Information of personnel roles and training requirements
References: Citations, user guides or other supporting SOPs

Standard Operating Procedures

Standard Operating Procedures



Quality Research Practices

Table 4: QPTR Core Checklist

Personnel Criteria

- Researcher has a training record on file ('Researcher' includes all technicians, students, postdocs, etc. Each needs to complete all of these points)
- Training record contains current CV
- Training record lists all completed trainings within appropriate timeframe
- Researcher has met with study design and biostatistics experts and has developed an approved study (see below)
- Researcher has demonstrated that documentation practices are satisfactory in both paper or online training
- Passed QA auditing of procedures
- For any techniques that require precise or tricky skills, researcher has mastered these skills exactly as stated in the SOP
- QA has agreed that this researcher has completed all required training both through OnBase as well as hands-on training as required
- No researcher on a study is currently 'not in good standing' with QA

Study Protocol Criteria

- The study protocol has been reviewed by a study design specialist and contains the following required elements:
  - Proper controls, adequate power, required gender balance of test subjects, blinding (feasible/required), randomization as necessary
- The methods planned are adequate to conduct the described hypothesis
- Methods for data analysis are outlined in advance, are feasible and have been established with other cores or services
- Plans exist for long term storage of both physical samples and data as needed
- The procedures in the study are contained in the list of approved procedures which can be conducted under the established quality system elements
- The relevant core has indicated current capacity to take on the outlined study
- Facilities listed in the study are all currently in good working order and have been cleared by QA as meeting the selected elements of the quality system
- All equipment to be used in this study has documented maintenance and calibration as appropriate

Table. Existing and New Strategies That *Annals* Uses to Guard against the Publication of Biased or Invalid Research

Strategy	Purpose
Prepublication	
Conflict disclosures	Alerts editors and reviewers to potential sources of bias
Peer review	Provides an opportunity for content experts to examine the work for potential threats to validity
Protocol submission	Enables editors to reconcile what researchers planned with what they report
Statistical review by an in-house team (may include requests for data and alternative analyses)	Enables independent statisticians to critique analysis and look for threats to validity
Trial registration	Permits public access to research plans; discourages suppression of unfavorable results
Publication	
Publication of conflicts of interest	Alerts readers to potential sources of bias
Publication of author contributions	Establishes accountability for specific components of the research process
Publication of detailed methods	Enables readers to critique methods and look for threats to validity
Postpublication	
Letters to the editor/rapid responses	Provides a venue for readers to air their concerns
New safeguards	
Publication of information about availability of protocol, statistical code, and data	Increases potential for reproducibility, allows greater scrutiny for potential threats to validity Permits confirmation of results by independent individuals

Laine C. Ann Internal Med 2007;146:450

Data collection, storage and analysis

Biomarkers

Types

RNA

DNA

Serum

Epigenetic

Proteomics

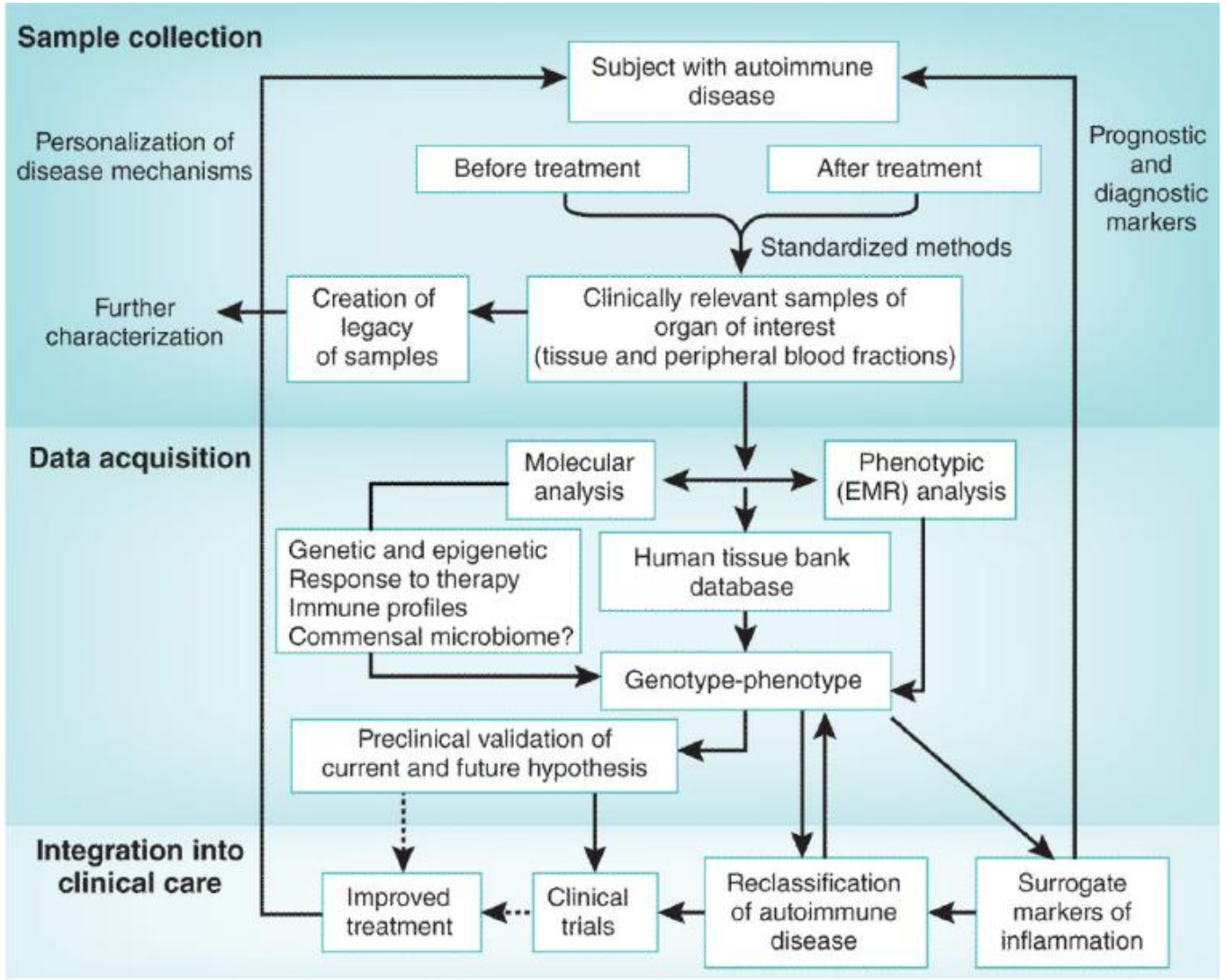
Single cell analysis

Cell populations-flow cytometry

Challenges

- Cost
- Storage
- Stability
- Validity
- Reproducibility

Samples in autoimmunity-related research



Blumberg RS. Nature Medicine 2012. 18(1):35

Biomarkers- Minimizing Bias and Error

Flow cytometry- unbiased gating and ongoing monitoring for batch effects

Double entry of data

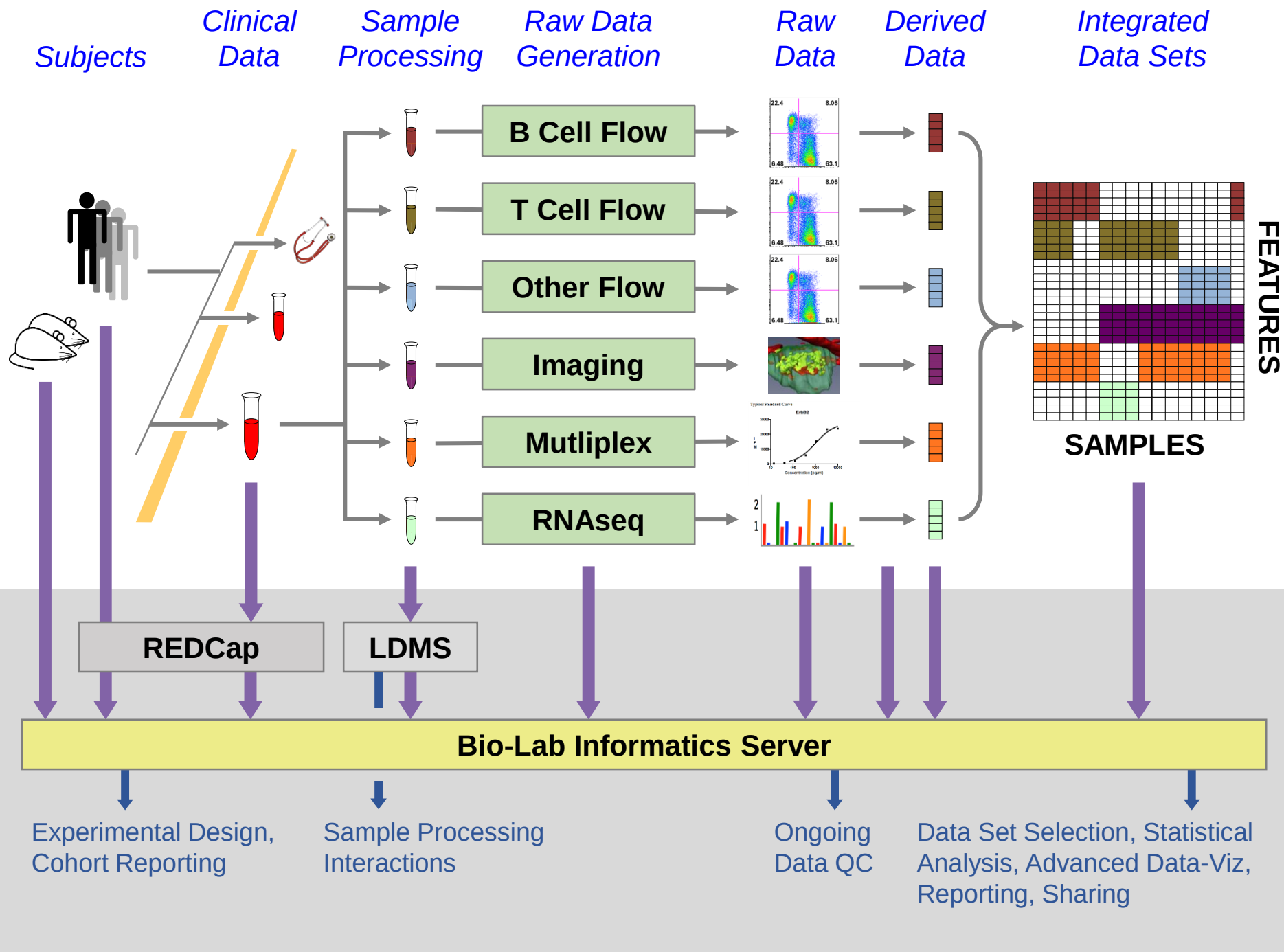
Close collaboration with biostatistics and bioinformatics

Ongoing training and education that centers on scientific rigor and reproducibility

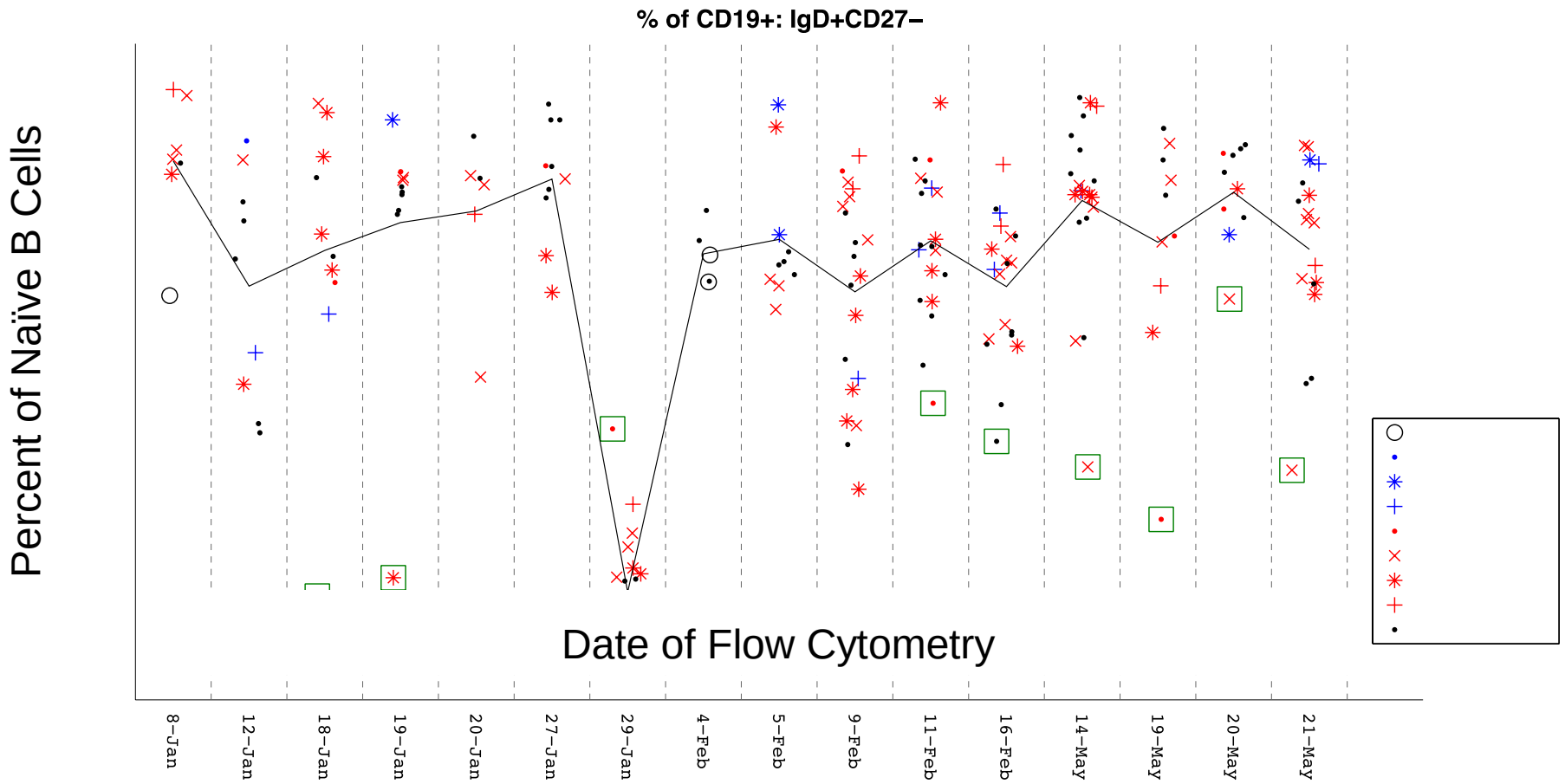
Integrated data capture

Data management is key to progress. Extensive cellular and molecular profiling of human subjects generates vast amounts of disparate data. Effective data management and integration solutions are essential to the preservation of this information in an interpretable form. Thus, data management efforts occurring 'behind the scenes' have an essential role to play in realizing the full potential of high throughput profiling approaches in human subjects.

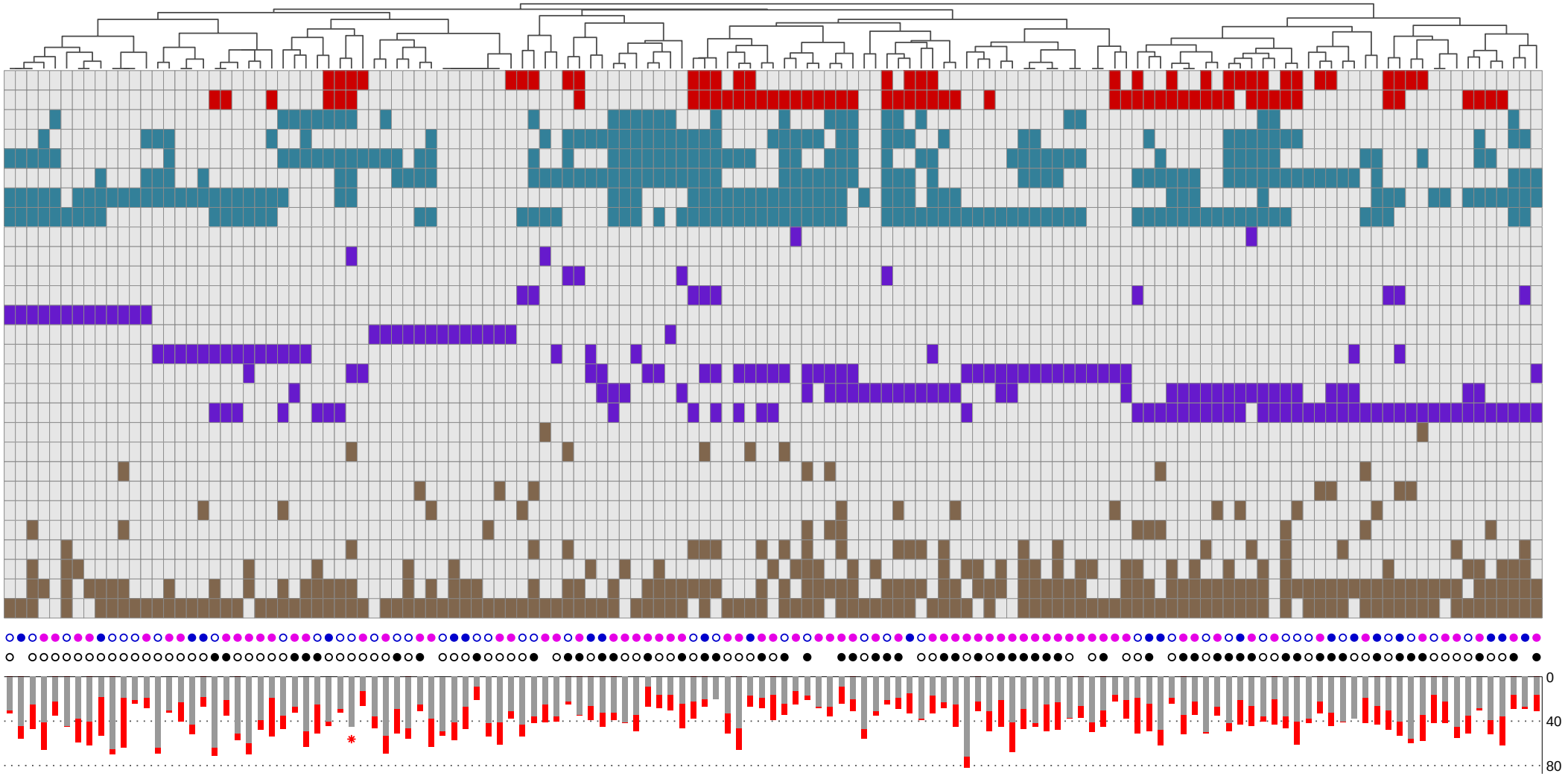
Chaussabel et al. (2010) Assessing the human immune system through blood transcriptomics. *BMC Biology* **8**:84 doi:10.1186/1741-7007-8-84


Subjects
Clinical Data
Sample Processing
Raw Data Generation
Raw Data
Derived Data
Integrated Data Sets
B Cell Flow
T Cell Flow
Other Flow
Imaging
Mutliplex
RNAseq
FEATURES
SAMPLES
REDCap
LDMS
Bio-Lab Informatics Server
Experimental Design, Cohort Reporting
Sample Processing Interactions
Ongoing Data QC
Data Set Selection, Statistical Analysis, Advanced Data-Viz, Reporting, Sharing

Detecting batch effects



Integrated data capture



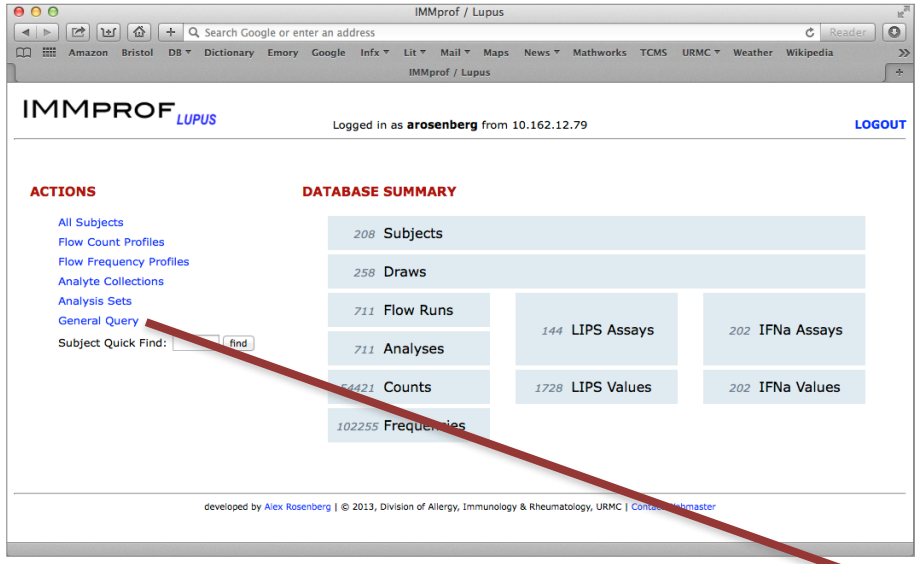
Disease Activity Measures

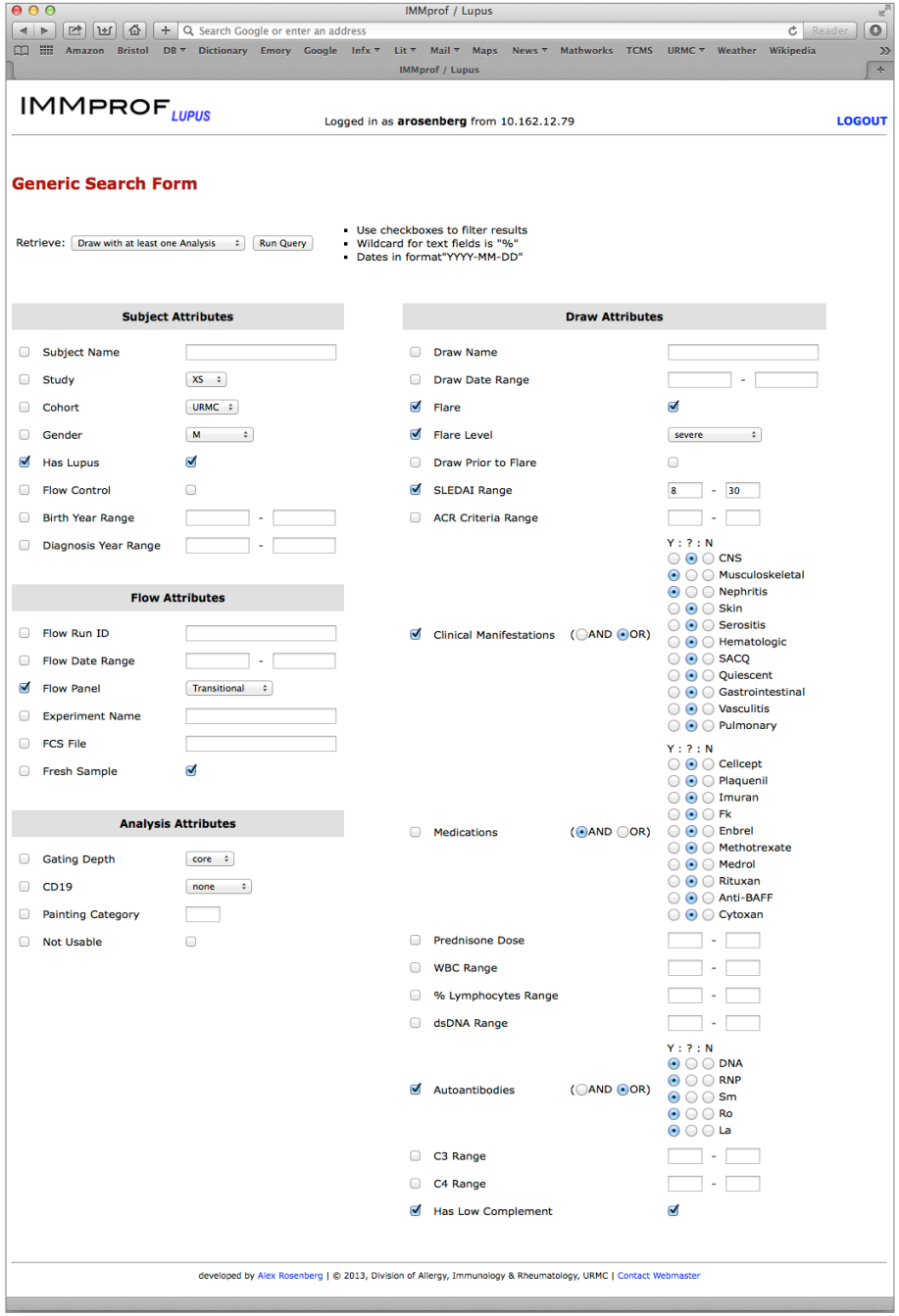
Labs

Clinical Assessment Variables

Meds

Demographics and Cohort Variables





"Show me samples that have associated analyzed flow cytometry data using the 'transitional B cell panel' that are derived from lupus patients with:

- severe flare
- SLEDAI ≥ 8
- musculoskeletal or nephritis manifestation
- any positive autoantibody measurement
- has low complement"

Take Home Points

Develop and maintain good research practices

Formally train and empower young investigators

Embrace big science

Formalize and make SOPS available to all

Integrated system to capture and store patient, imaging, biomarker data

Recruit biostatisticians and bioinformatics specialists to your team

Resources

- NIH website
- NIH modules on scientific rigor and reprocibility
- ARRIVE guidelines

- *Resnik, DB. Reproducibility and research integrity. Account. Res.2017;24(2):122*