

From biobanking to bioinformatics

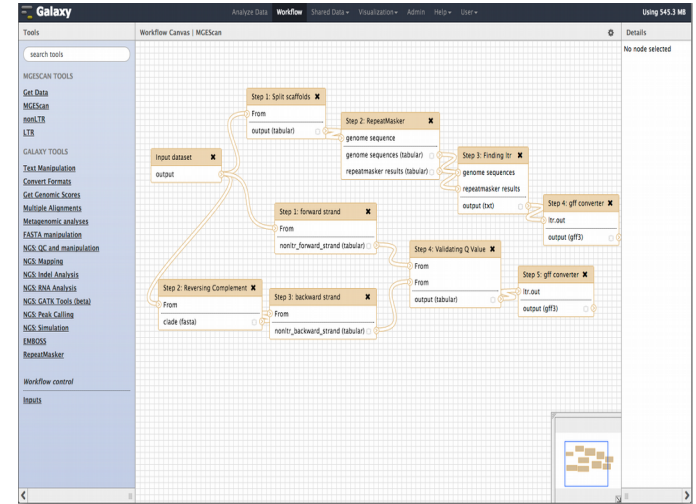
Hocine Bendou
SANBI - UWC April 2018



Introduction



Fastq
Files



Pre-analytical Phase

Analytical Phase

Introduction

- > 53% irreproducibility in clinical research
- > 75% in pharmaceutical companies
- > 28 billion \$ a year wasted in US alone



PERSPECTIVE

The Economics of Reproducibility in Preclinical Research

Leonard P. Freedman^{1*}, Iain M. Cockburn², Timothy S. Simcoe^{2,3}

1 Global Biological Standards Institute, Washington, D.C., United States of America, **2** Boston University School of Management, Boston, Massachusetts, United States of America, **3** Council of Economic Advisers, Washington, D.C., United States of America

Introduction

Clinical Chemistry 61:7
914-934 (2015)

Review

Preanalytical Variables Affecting the Integrity of Human Biospecimens in Biobanking

Christina Ellervik^{1,2,3*} and Jim Vaught^{4,5}

British Medical Bulletin, 2015, **114**:29–38

doi: 10.1093/bmb/ldv012

Advance Access Publication Date: 24 April 2015



Developments in biospecimen research

Jim Vaught*

International Society for Biological and Environmental Repositories, Kensington, MD, USA

*Correspondence address. International Society for Biological and Environmental Repositories, 3405 Wake Drive, Kensington, MD 20895, USA. E-mail: jvaught44@gmail.com

Accepted 12 March 2015

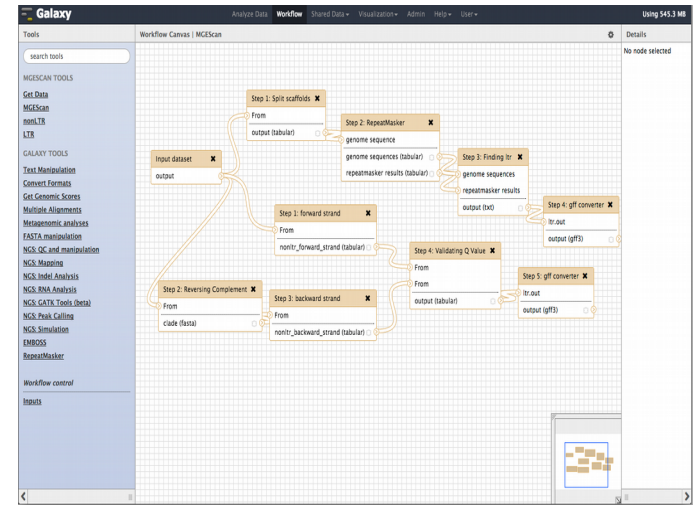
Introduction



...



**Fastq
Files**



Results

Biobank



CKB sample storage in China

- Facility for collection, preservation, storage and supply of biological samples (whole blood, DNA, RNA, ...) and associated data (genetic, phenotypes, environmental, ...)
- Follows guidelines and standardised operating procedures (SOPs)
- Ex:
 - Avoid thaw/freeze cycle
 - Security; store samples in different locations

Laboratory Information Management System (LIMS)

- Manage the operations within a biobank
- Covers the life cycle of a biospecimen from collection to destruction
- Ex: Baobab LIMS



Laboratory Information Management System (LIMS)

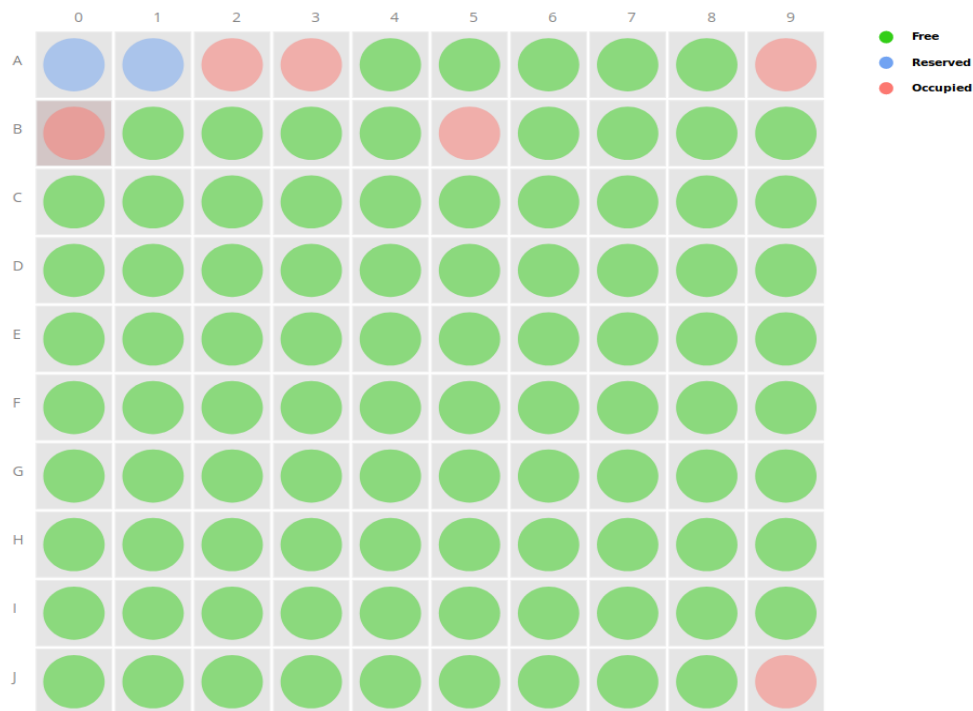


Occupancy

Total Positions	100
Available	92
Reserved	2
Occupied	6

Freezerfarm-01.freezer-001.Partition-01.Racks-01.Tray-001.Box-001.11

Sample ID	DNA-0001
Sample Name	DNA-0001
Volume (ul)	90.00
Subject ID	CNV967.1BAB



Find the location of the biospecimen with barcode = **DNA-001**

Biospecimen sharing



...

API, CSV



H3Africa
Human Heredity & Health in Africa

Data & Biospecimen Catalogue

[Catalogue](#) [About](#) [Contact](#)

Search

Catalogue Content
Last release Apr 26, 2017

Studies	5
Datasets	106
Biospecimens	106

Advanced Search

Study Participant Biospecimen

Study Name

Study Design

Disease

☐ Only studies with Biospecimens

☐ Only studies with datasets

SEARCH

Available Studies:

Collaborative African Genomics Network [more...](#)

Acronym: [CalGEN](#)

Design: Case-Control

Specimens with aliquots: 85

Participants with data: 89

[VIEW](#) [ADD](#)

The Genomics of Schizophrenia in the South African Xhosa Population [more...](#)

Acronym: [SAX](#)

Design: Case-Control

Specimens with aliquots: 4

Participants with data: 4

[VIEW](#) [ADD](#)

H3Africa Diabetes Study [more...](#)

Acronym: [H3AD](#)

Design: Case-Control

Specimens with aliquots: 4

Participants with data: 2

[VIEW](#) [ADD](#)

Stroke Investigative Research & Educational Network [more...](#)

Acronym: [SIREN](#)

Design: Case-Control

Specimens with aliquots: 17

Participants with data: 17

[VIEW](#) [ADD](#)

An integrated approach to the identification of genetic determinants of susceptibility to trypanosomiasis [more...](#)

Acronym: [TrypanoGEN](#)

Design: GWAS

Specimens with aliquots: 0

Participants with data: 0

[VIEW](#) [ADD](#)

H3ABioNet
Pan African Bioinformatics Network for H3Africa

Help
Contact

Source
Contribute

Made by SANBI-CBIO Team

Data Biospecimen Access Committee

8. **Consent and Approvals** Please confirm that you have obtained all approvals required for this project by the rules and regulations of your jurisdiction, including your institution's institutional rules and any relevant national rules, by ticking the following box: (you may be requested to provide a copy of the consent form, particularly for biospecimen requests)

- ☐ Yes, I obtained all approvals required
- ☐ No

For Biospecimen requests

I have read and agree to abide by the terms and conditions outlined in the H3Africa Biospecimen Sharing Policy and will sign the Material Transfer Agreement negotiated with the host biorepository

- ☐ Yes
- ☐ No

For Data requests

I have read and agree to abide by the terms and conditions outlined in the Data Access Agreement and the Publication Policy?

- ☐ Yes
- ☐ No

9. Please upload your query csv file (this is required)

QUERY

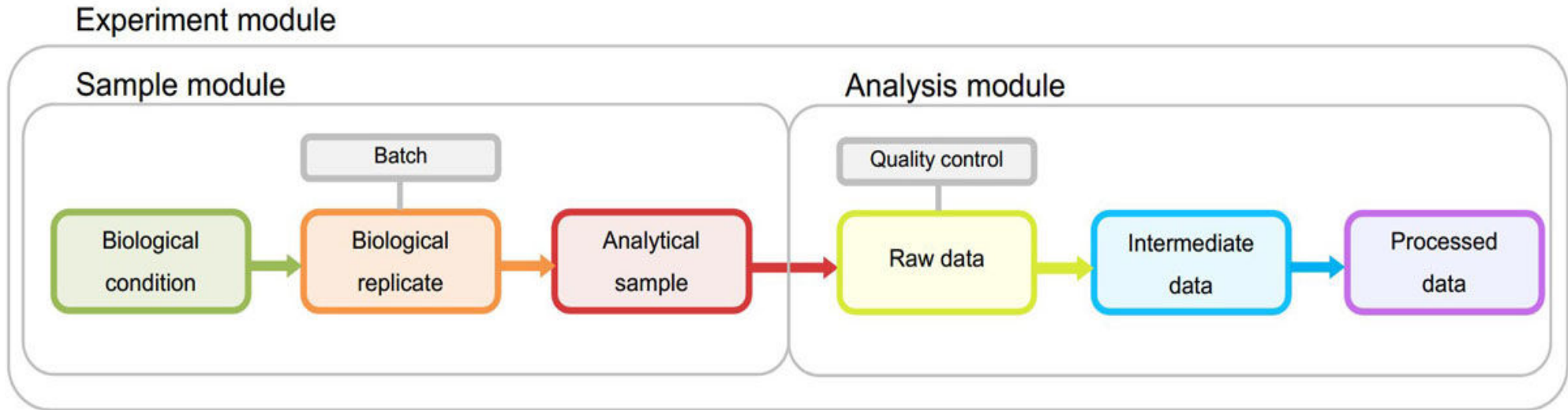
Upload a csv file

REQUEST ➤

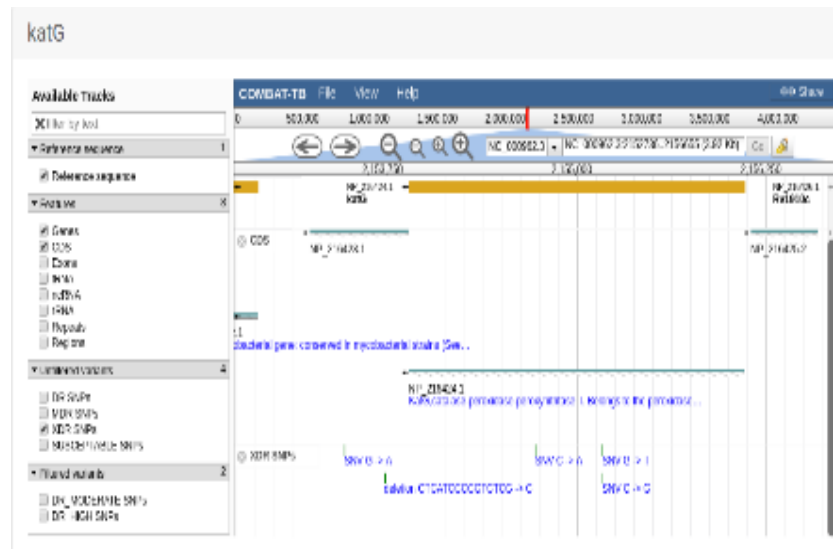
CANCEL ✕

Experiment Management System (EMS)

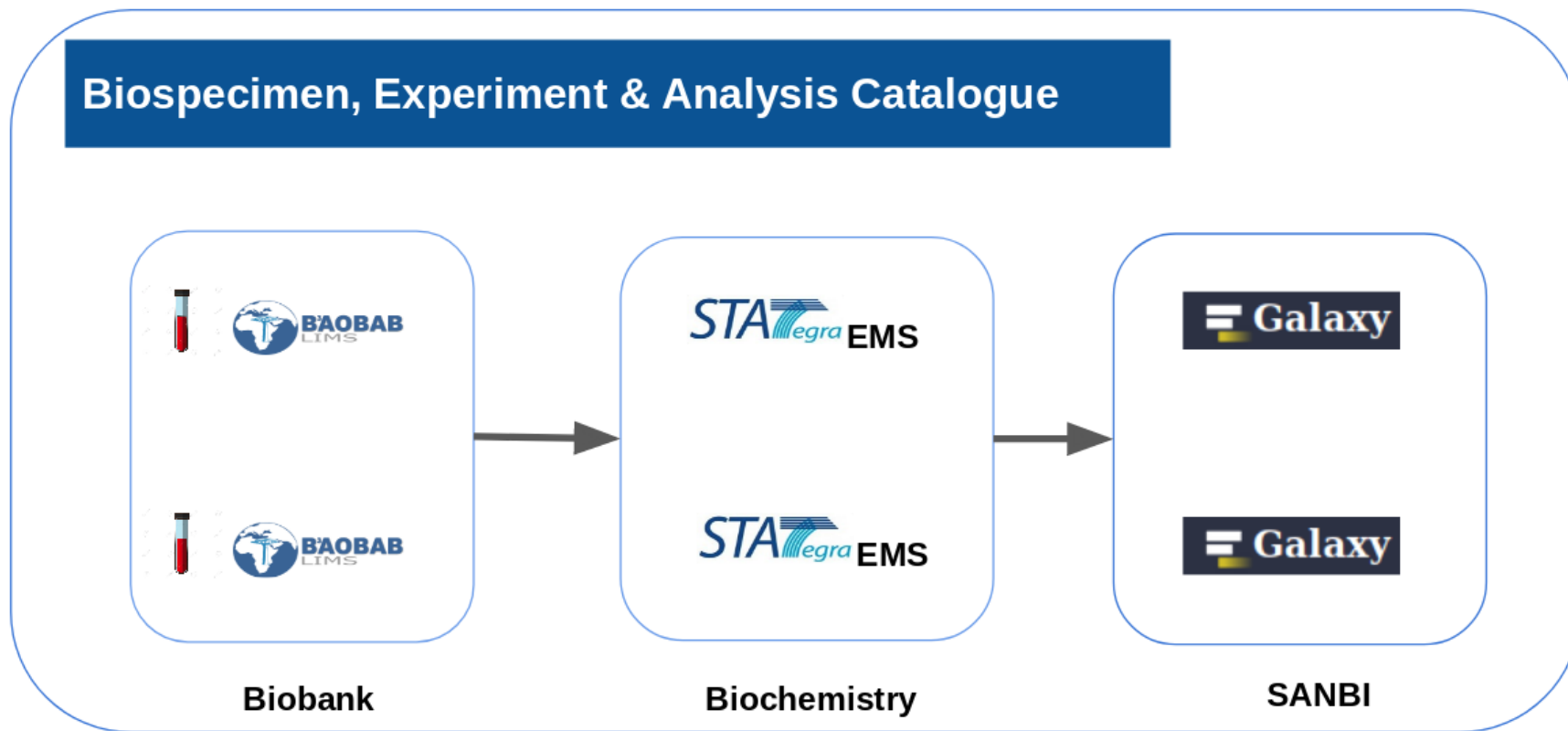
Stategra EMS



The screenshot shows the Galaxy workflow editor interface. On the left, there is a sidebar with navigation links: 'Galaxy', 'Tools', 'MGESCAN TOOLS', 'Get Data', 'MGEScan', 'nonLTR', 'LTR', 'GALAXY TOOLS', 'Text Manipulation', 'Convert Formats', 'Get Genomic Scores', 'Multiple Alignments', 'Metagenomic analyses', 'FASTA manipulation', 'NGS: QC and manipulation', 'NGS: Mapping', 'NGS: Indel Analysis', 'NGS: RNA Analysis', 'NGS: GATK Tools (beta)', 'NGS: Peak Calling', 'NGS: Simulation', 'EMBOSS', 'RepeatMasker', 'Workflow control', and 'Inputs'. The main area displays a workflow canvas with a grid background. The workflow consists of several steps: 'Input dataset' (orange box) connects to 'Step 1: Split scaffolds' (orange box) and 'Step 1: forward strand' (orange box). 'Step 1: Split scaffolds' connects to 'Step 2: Repeating Masker' (orange box). 'Step 1: forward strand' connects to 'Step 2: Reversing Complement' (orange box). 'Step 2: Repeating Masker' connects to 'Step 3: Finding ltr' (orange box). 'Step 2: Reversing Complement' connects to 'Step 3: Finding ltr' (orange box). 'Step 3: Finding ltr' connects to 'Step 4: Validating Q Value' (orange box). 'Step 4: Validating Q Value' connects to 'Step 5: gff converter' (orange box). The final outputs are 'ltr.out' and 'output (gff3)' (orange boxes). The right sidebar shows 'Details' and 'No node selected'.



Final workflow



Conclusion



African Baobab tree



Thank you

