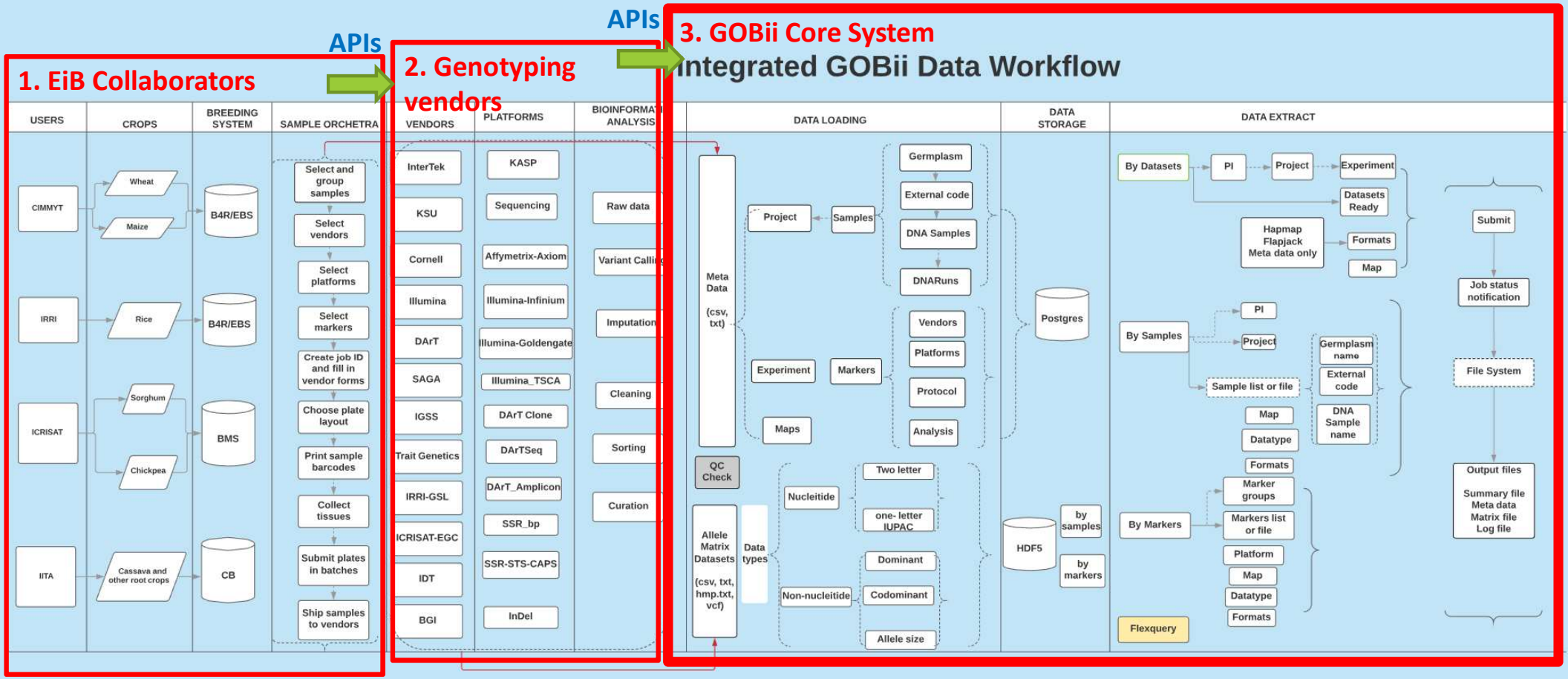
Integrated Genomic Selection Galaxy Analysis Pipeline and Workflows

Star Yanxin Gao
yg28@cornell.edu
http://gobiiproject.org

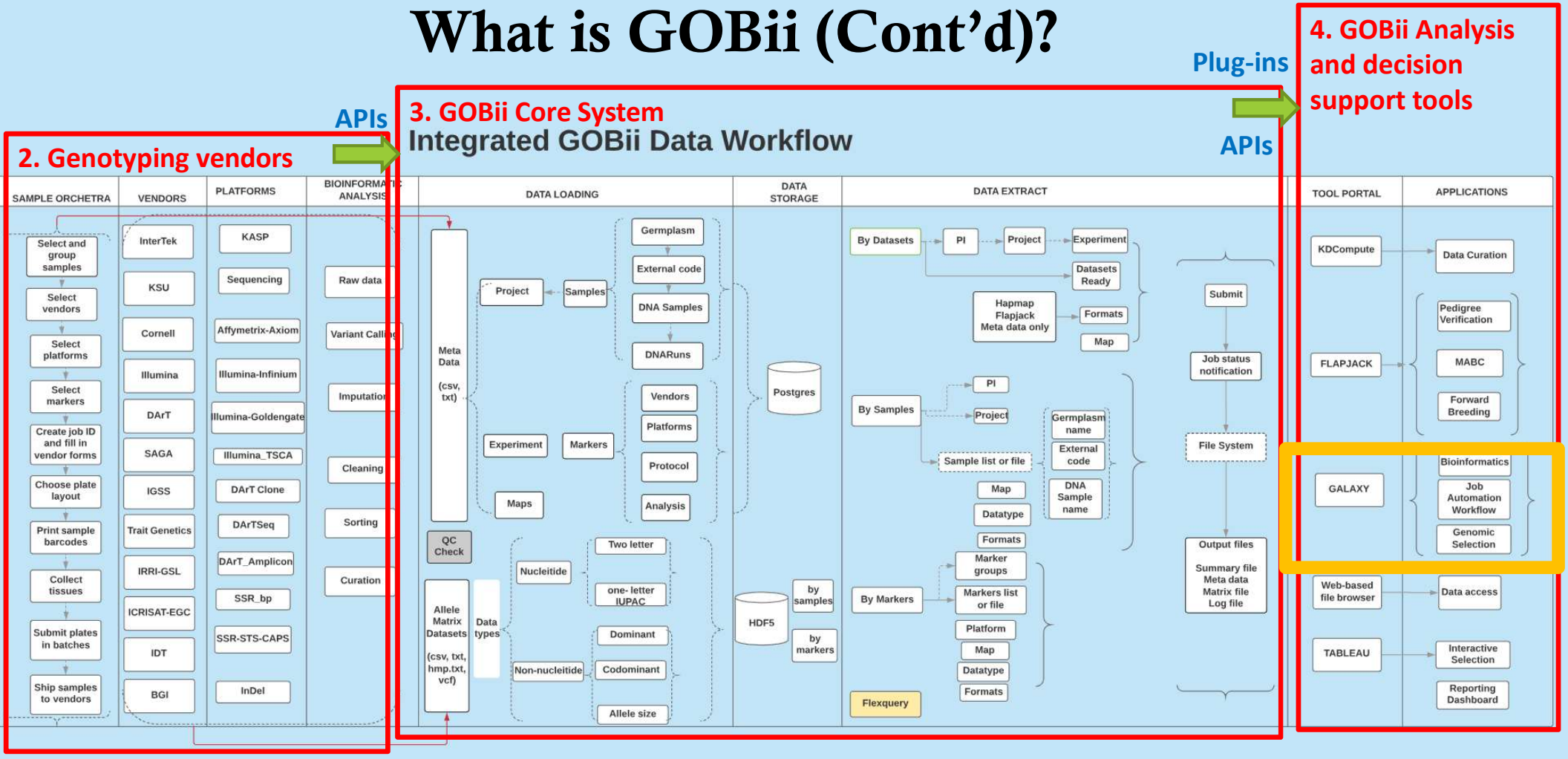
GOBii: A Global Team

http://gobiiproject.org

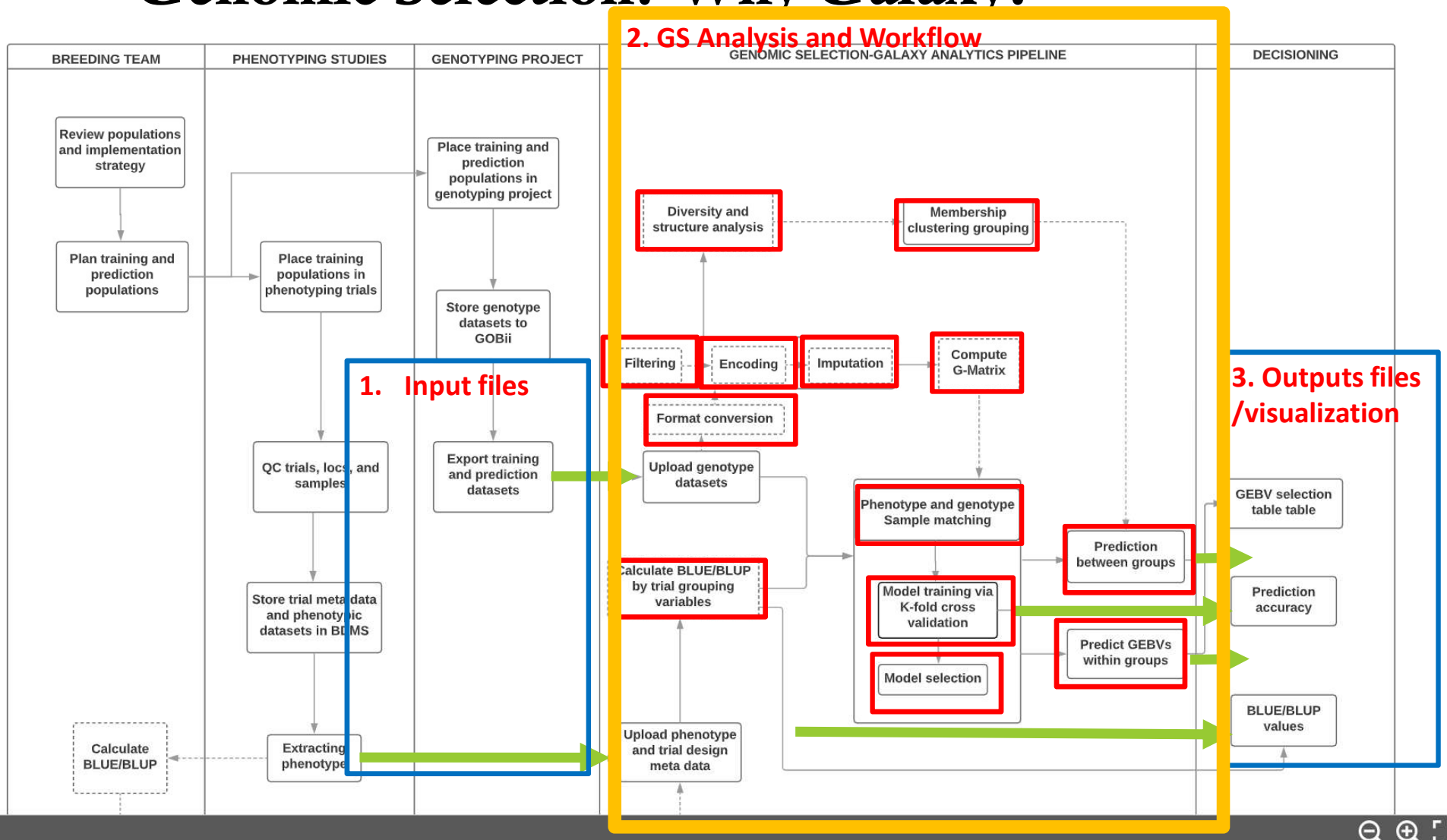
What is GOBii?



What is GOBii (Cont'd)?



Genomic Selection? Why Galaxy?



Importing Workflows and Test Datasets

This first workflow is using Beagle for Imputation. It connects BLUP calculator, Imputation and GEBV prediction.

Galaxy Workflow | Genomic selection (workflow1)

Input Datasets

Genotyping dataset in hapmap format

Import datasets

Galaxy Dataset | test2.hmp

Phenotyping dataset

Galaxy Dataset | Phenotype_test_file_1015 GS-Galaxy.tab

Complete history for workflow1

Galaxy History | Genomic_selection_workflow1

Your workflows

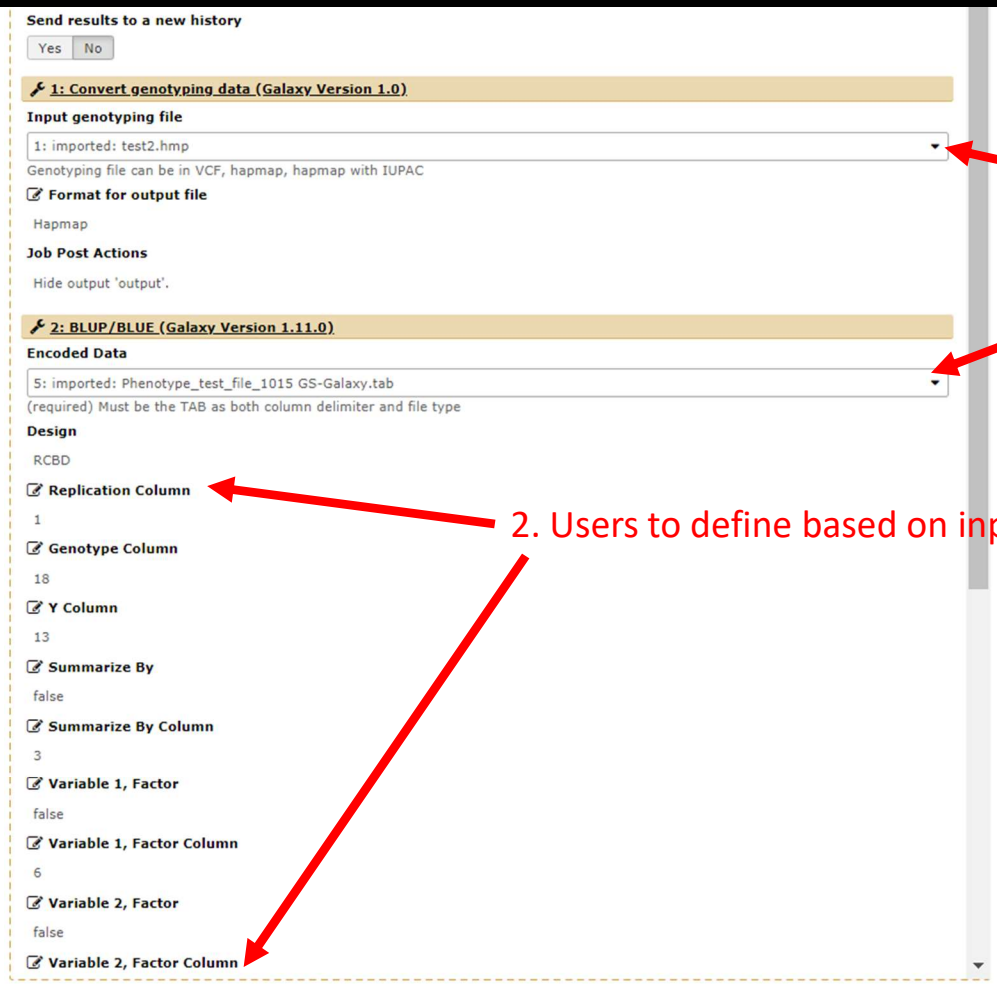
Run workflow

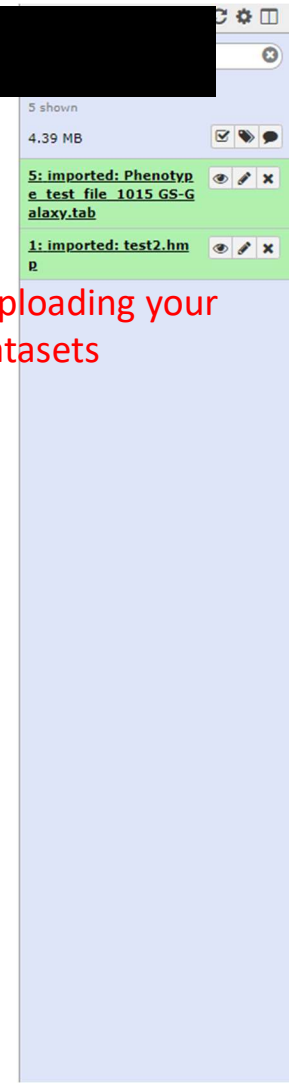
search for workflow...

Name	Tags	Owner	# of Steps	Published	Show in tools panel
SNiPlay diversity workflow		You	17	Yes	
Genomic selection (workflow1)		You	8	Yes	
G) - Flapjack Analysis		You	4	No	

Edit
Run
Share
Download
Copy
Rename
View
Delete

Define datasets and parameters





1. Users select input files: uploading your own or importing shared datasets

2. Users to define based on input file

Genomic Selection (GS) demo-workflow1

Galaxy / EiB-demo | Analyze Data | Workflow | Shared Data | Visualization | Help | User

Published Pages | dereeper | Genomic Selection Pipeline1

GOBii Genomic Selection Galaxy Workflows

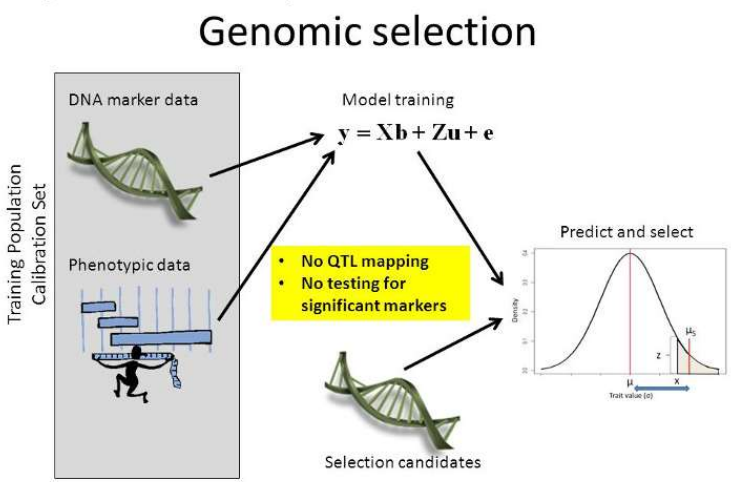
Pipeline 1: Prediction of genomic breeding values (GEBV)

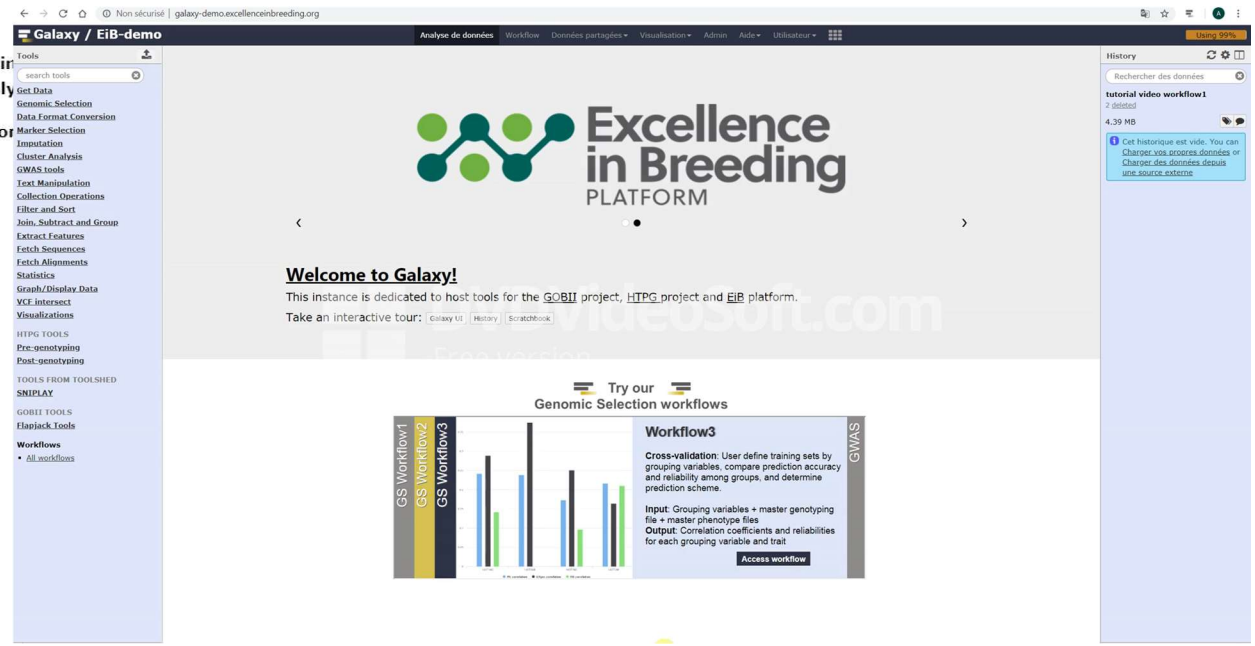
Prediction of genomic estimated breeding values (GEBV) for untested individuals within

Calculate the genomic estimated breeding values (GEBVs) for individuals that have only training dataset.

If sub-structure or groups exist within the test dataset, run Workflows 2 or 3 first befor

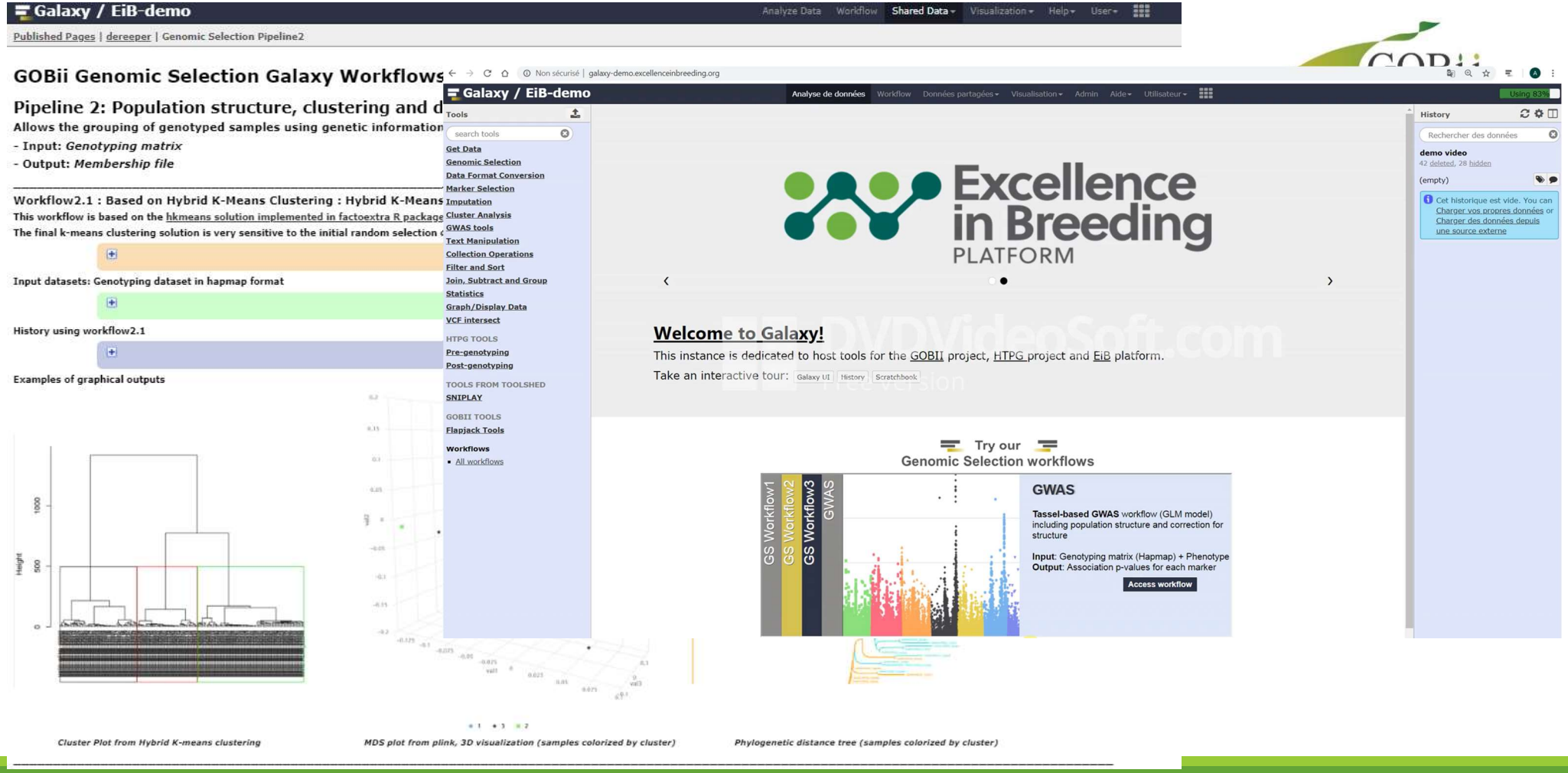
- Input: *Genotyping matrix + Phenotyping file*
- Output: *Table of observed and predicted values for each trait*



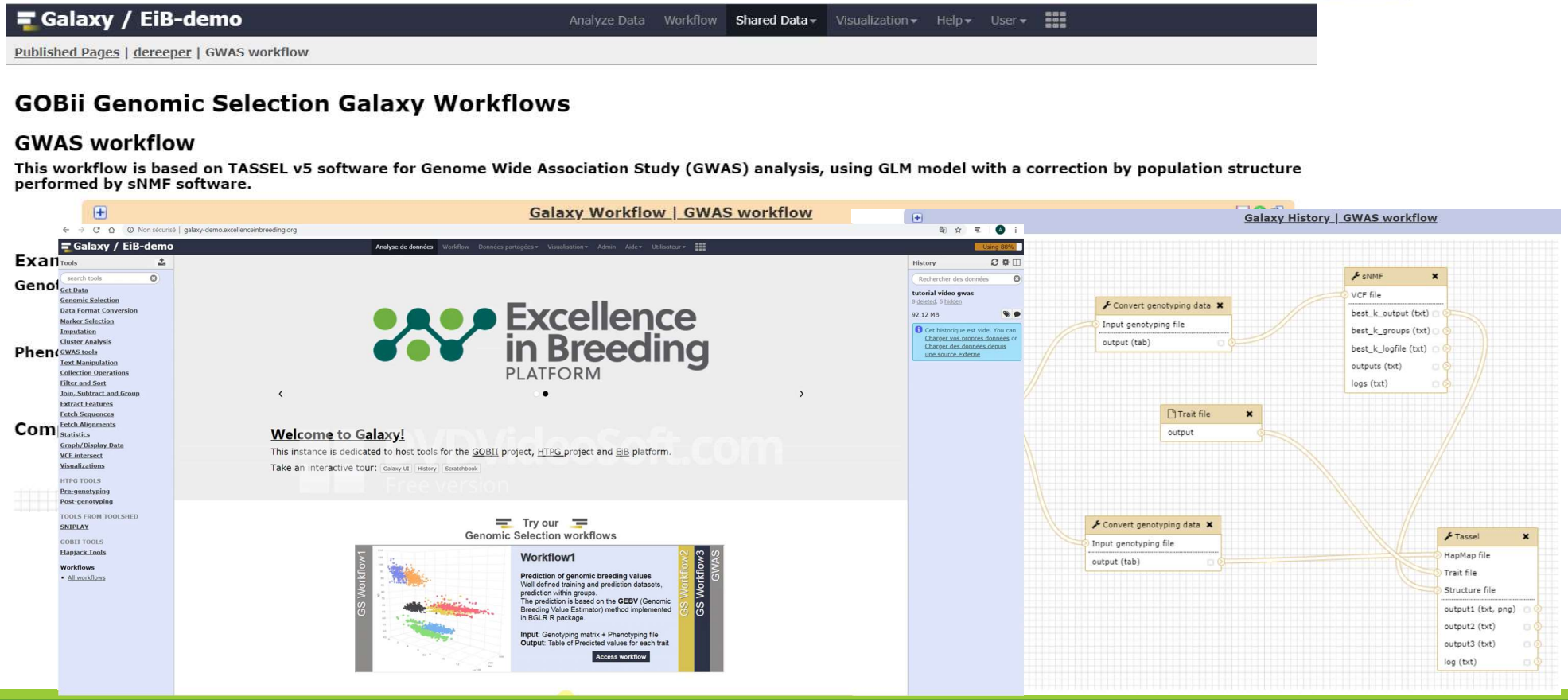


C. Kadam, D & J. Lorenz, A. (2018) Toward Redesigning Hybrid Maize Breeding Through Genomics-Assisted Breeding

Genomic Selection (GS) demo-workflow 2



Genome Wide Association Analysis workflow demo (GWAS)



GOBii GS-Galaxy pipeline

http://galaxy-demo.excellenceinbreeding.org/

- Free and open to the public
- Teaching and training platform to the next generation molecular breeders
- GS testing team collaboration playground to share datasets, history, workflows, and parameters
- Data exploration for samples genotyped but not phenotyped vice versa and study planning- **Sample matching**
- Handling bioinformatics and file **conversions-encoding, filtering, format conversion**
- Phenotype analysis and visualization- **BLUP/BLUE calculator**
- Batch analysis with reproducible parameters for multiple traits and variables, factors- **BLUP/BLUE and GEBV calculators**
- Sub-structure analysis- **Clustering and diversity analysis tools**
- Well established customer-defined GS and GWAS workflows and training methods- **Workflows/Pipelines**

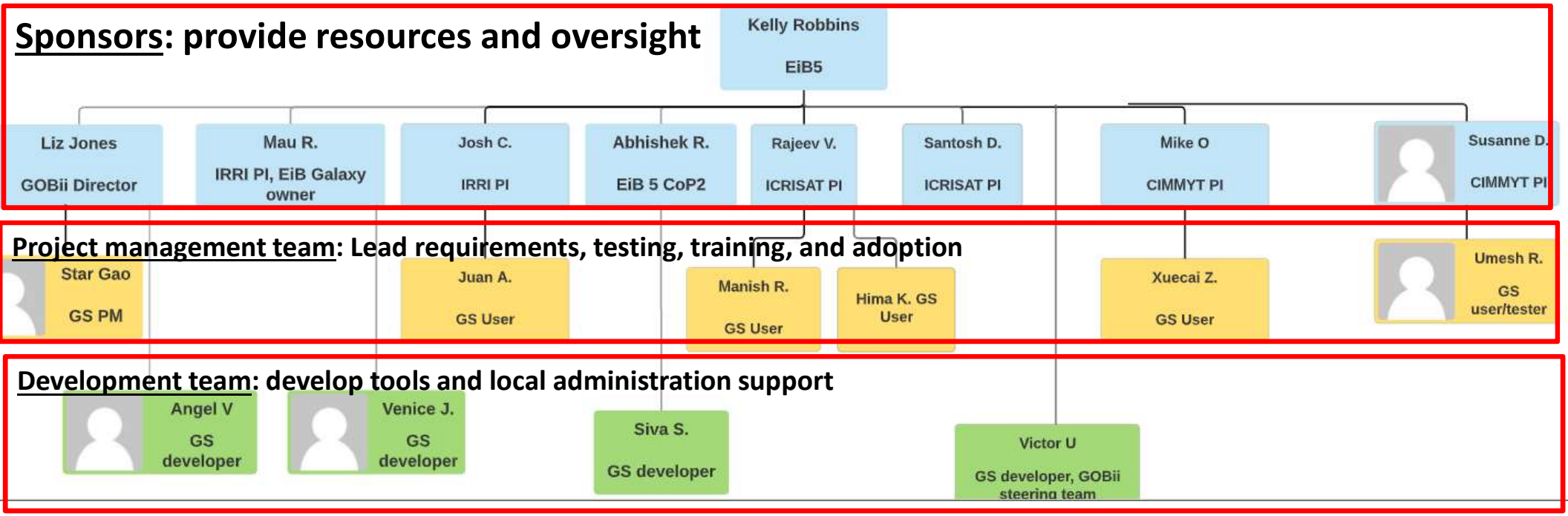
GOBii GS-Galaxy pipeline

http://galaxy-demo.excellenceinbreeding.org/

ARE NOT

- GS statisticians' model development, who are proficient with command-line implementation
- Free testing site of large datasets-need to download tools and workflows onto your own server
- For producing publication without a sound understanding or data interpretation

Acknowledgements



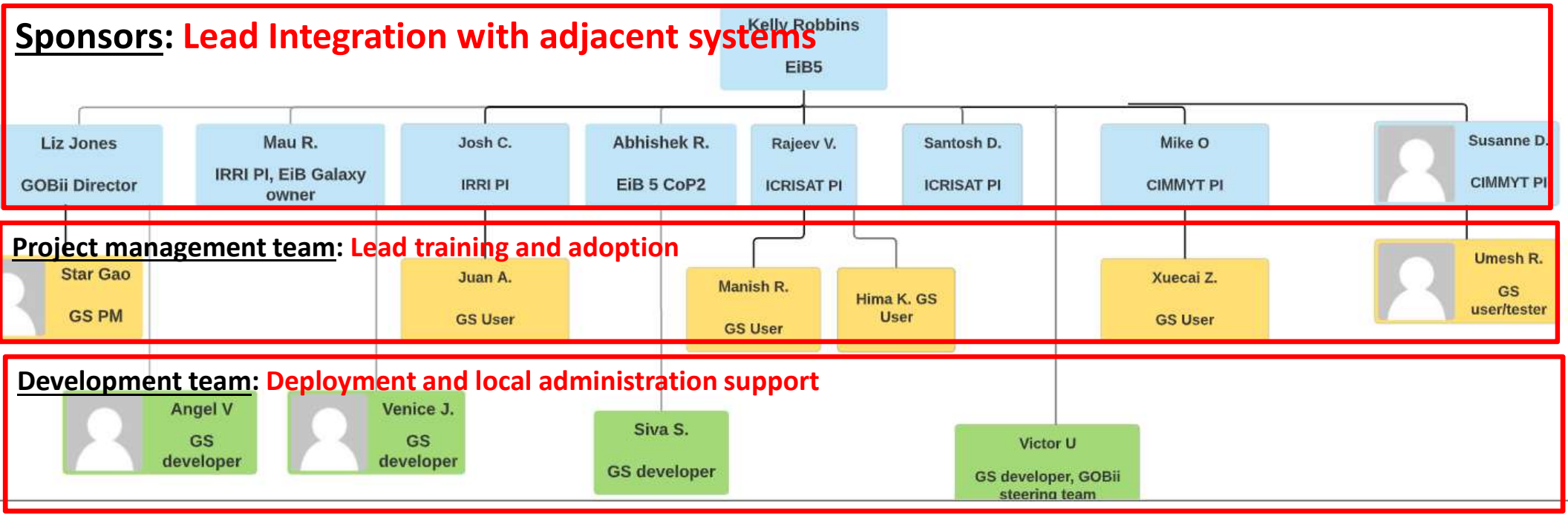
EiB contractor: optimize and integrate tools, set up workflows and visualization, enable toolshed and system admin training

Alexis DEREEPER

Contributors:

Paulino Perez, Pancho Jose Crossa, Fernado Toledo, Juan BURGUEÑO, Jessica Rutkoski, Demya Chebotarov, Jean-Luc Jannink, Isaak Yosief Tecle, Nicholas Santantonio, **Kate Dreher**, Claudio Ayala, Félix SanVicente, Mark Sorrells, **Yoseph Beyene**, Manje Gowda, Yaw Nti-Adde

What's next?



EiB contractor(s): Toolshed, server, and system admin training
Alexis DEREEPER

Community-based development and support

Contact us at *http://gobiiproject.org for collaboration, training, and customization*