

Search

- Search for valid tools

- Search for workflows

Valid Galaxy Utilities

- Tools

- Custom datatypes

- Repository dependency definitions

- Tool dependency definitions

All Repositories

- Browse by category

Repositories I Can Change

- Repositories I own

- Reviewed repositories I own

- Deprecated repositories I own

- My writable repositories

- Reset metadata on my repositories

- Latest revision: missing tool tests

- Latest revision: installation errors

- Latest revision: failing tool tests

- Latest revision: skip tool tests

- Latest revision: all tool tests pass

- Latest revision: invalid tools

Available Actions

- Create new repository

- Import repository capsule

Reviewing Repositories

- Repositories ready for review

- All repositories with no reviews

- All reviewed repositories

- Manage review components

Reviewing Repositories With Tools

- Latest revision: missing tool tests

- Latest revision: installation errors

- Latest revision: failing tool tests

- Latest revision: skip tool tests

- Latest revision: all tool tests pass

- Latest revision: invalid tools

Categories

Name	Description	Repositories
Assembly	Tools for working with assemblies	28
ChIP-seq	Tools for analyzing and manipulating ChIP-seq data.	4
Computational chemistry	Tools for use in computational chemistry	18
Data Source	Tools for retrieving data from external data sources	17
Genome-Wide Association Study	Utilities to support Genome-wide association studies	1
Imaging	Utilities to support imaging	3
Metabolomics	Tools for performing metabolomics	2
Metagenomics	Tools for performing metagenomics	14
Micro-array Analysis	Tools for performing micro-array analysis	9
Next Gen Mappers	Tools for the analysis and handling of Next Gen sequencing data	56
Ontology Manipulation	Tools for manipulating ontologies	7
Phylogenetics	Tools for performing phylogenetic analysis	7
Proteomics	Tools enabling the study of proteins	49
RNA	Utilities for RNA	11
Sequence Analysis	Tools for performing Protein and DNA/RNA analysis	175
Statistics	Tools for general statistics	48
Systems Biology	Systems biology	6
Text Manipulation	Tools for manipulating data	34
Tool Dependency Packages	Repositories that contain tool dependency packages	166
Tool Generators	Tools that make or help make new tools	1
Transcriptomics	Tools for use in the study of Transcriptomics.	9

Introduction to the Tool Shed: Automated tool installation and robust dependency control for reproducible analysis

Galaxy Australasia Workshop

March 2014

Ross Lazarus

Outline

- Galaxy tools
- Tool Shed: VCS for tools
- Strict dependency control
- Low impedance distribution
- Demonstration
- Commodity reproducible analysis

Galaxy: tools, data and results

Galaxy / @BakerIDI

Analyze Data Workflow Shared Data Visualization Admin Help User

Using 57%

Tools Load Data

search tools

BakerIDI

SR Test/Repair/BWA Tools

Local unreliable SR Quality Tools

Get Data

Send Data

Repeats and Complexity

Lift-Over

Text Manipulation

Filter and Sort

Join, Subtract and Group

Convert Formats

Extract Features

Fetch Sequences

Fetch Alignments

Get Genomic Scores

Operate on Genomic Intervals

Statistics

Wavelet Analysis

Graph/Display Data

Regional Variation

Multiple regression

Multivariate Analysis

Evolution

Motif Tools

Multiple Alignments

Metagenomic analyses

FASTA manipulation

NGS: QC and manipulation

Phenotype Association

NGS: Assembly

NGS: Mapping

NGS: RNA Analysis

NGS: Pivcoteo

NGS: Pivcos

NGS: SAM Tools

NGS: GATK Tools (beta)

NGS: Peak Calling

NGS: Simulation

VCF Tools

Epigenetics

Picard

SnEff 3.3

Table: GSEA Results Summary

Dataset	SAFHSgugagetoTabletabular
Phenotype	NoPhenotypeAvailable
Upregulated in class	na_pos
GeneSet	BASSO_HAIRY_CELL_LEUKEMIA_DN
Enrichment Score (ES)	0.8217315
Normalized Enrichment Score (NES)	2.1532662
Nominal p-value	0.0
FDR q-value	0.0
FWER p-Value	0.0

Enrichment plot:
BASSO_HAIRY_CELL_LEUKEMIA_DN

Fig 1: Enrichment plot: BASSO_HAIRY_CELL_LEUKEMIA_DN
Profile of the Running ES Score & Positions of GeneSet Members on the Rank Ordered List

Table: GSEA details [plain text format]

	PROBE	GENE SYMBOL	GENE_TITLE	RANK IN GENE LIST	RANK METRIC SCORE	RUNNING ES	CORE ENRICHMENT
1	IGFBP3			0	74.686	0.3951	Yes
2	MAF			65	15.338	0.4733	Yes

History

safhs march 2014
628.7 MB

17: age_SAFHS_MSIG
DB_c2c3c5_gsea.html

16: hdl_SAFHS_MSIG
DB_c2c3c5noca_gsea.html

15: male_SAFHS_MSIG
DB_c2c3c5_gsea.html

12: male_SAFHS_ENC
ODE_TFBS_gsea.html

11: hdl_SAFHS_ENC
DE_TFBS_gsea.html

10: SAFHSgugageto
pTable.tabular

9: SAFHSgugageto
pTable.tabular

8: SAFHS_male_topTa
ble.xls

7: SAFHS_hdl_topTabl
e.xls

6: age_SAFHS_ENC
ODE_TFBS_gsea.html

5: age_SAFHS_gsea.ht
ml

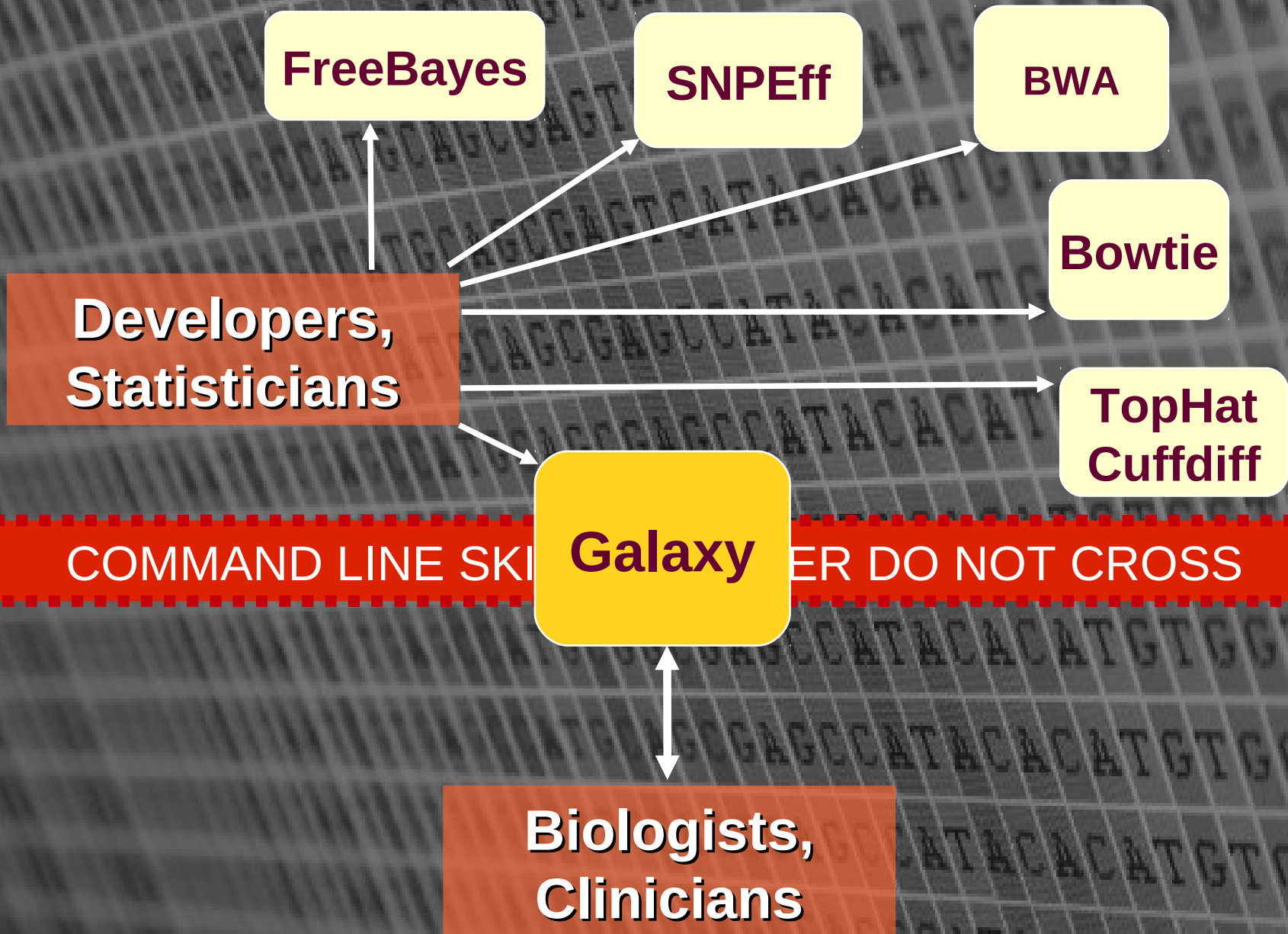
4: SAFHSgugageto
pTable.tabular

1: SAFHS_age_topTabl
e.xls

Galaxy Framework is tool agnostic

- Framework: focus of dev team
- Galaxy = UI, histories, viz, data, jobs..
- So, new component
- Tool Shed: tool server(s) for Galaxy
- Galaxy and tool distribution ***decoupled***

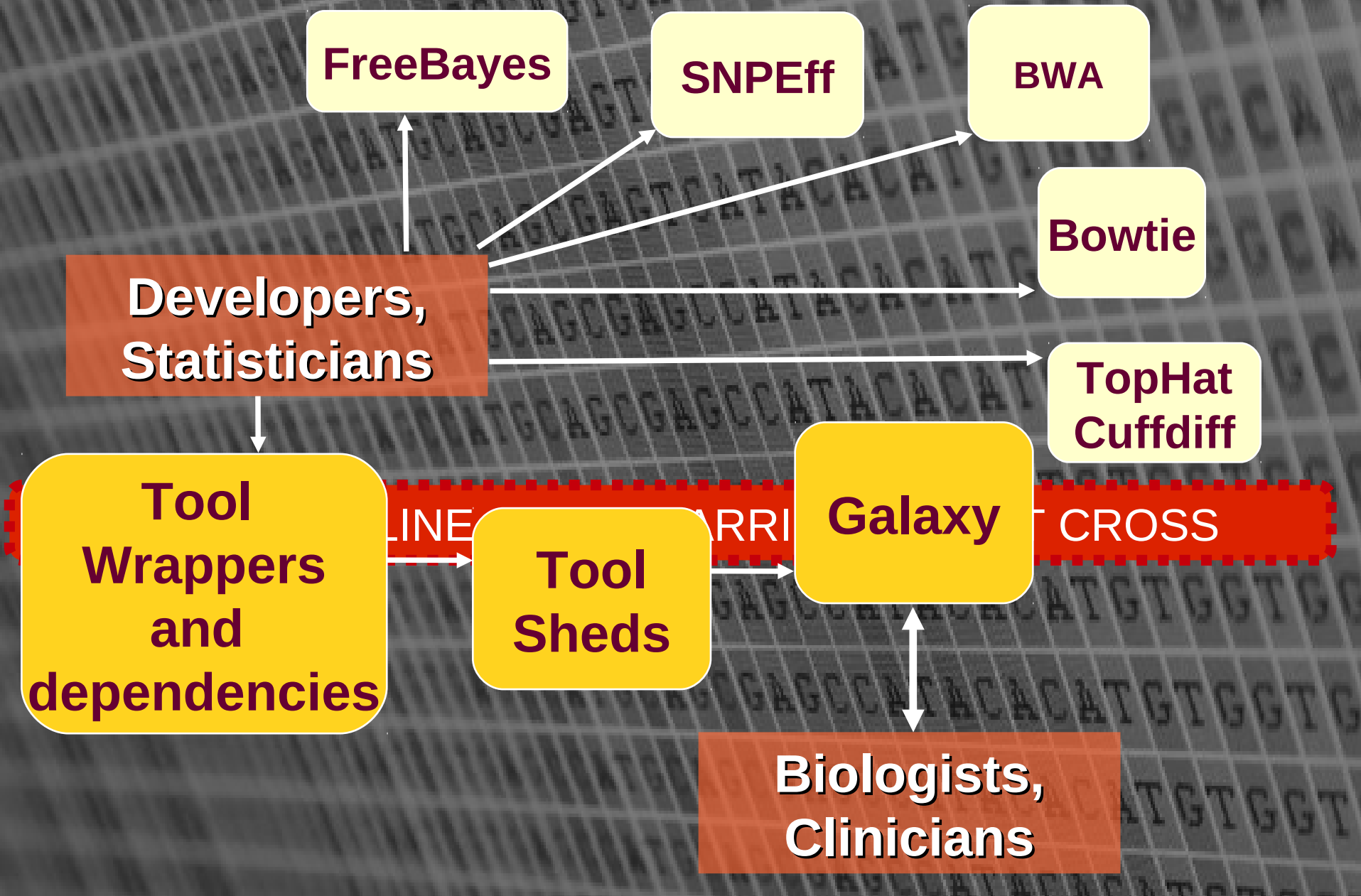
Where does Galaxy fit?



Historical tool distribution model

- Galaxy distributions: code + tool wrappers
- Executables always from local system
- eg: bwa wrapper → system “bwa”
- If system updated, redo → bad! different!
- Weakened reproducibility claim

Where does Galaxy fit?



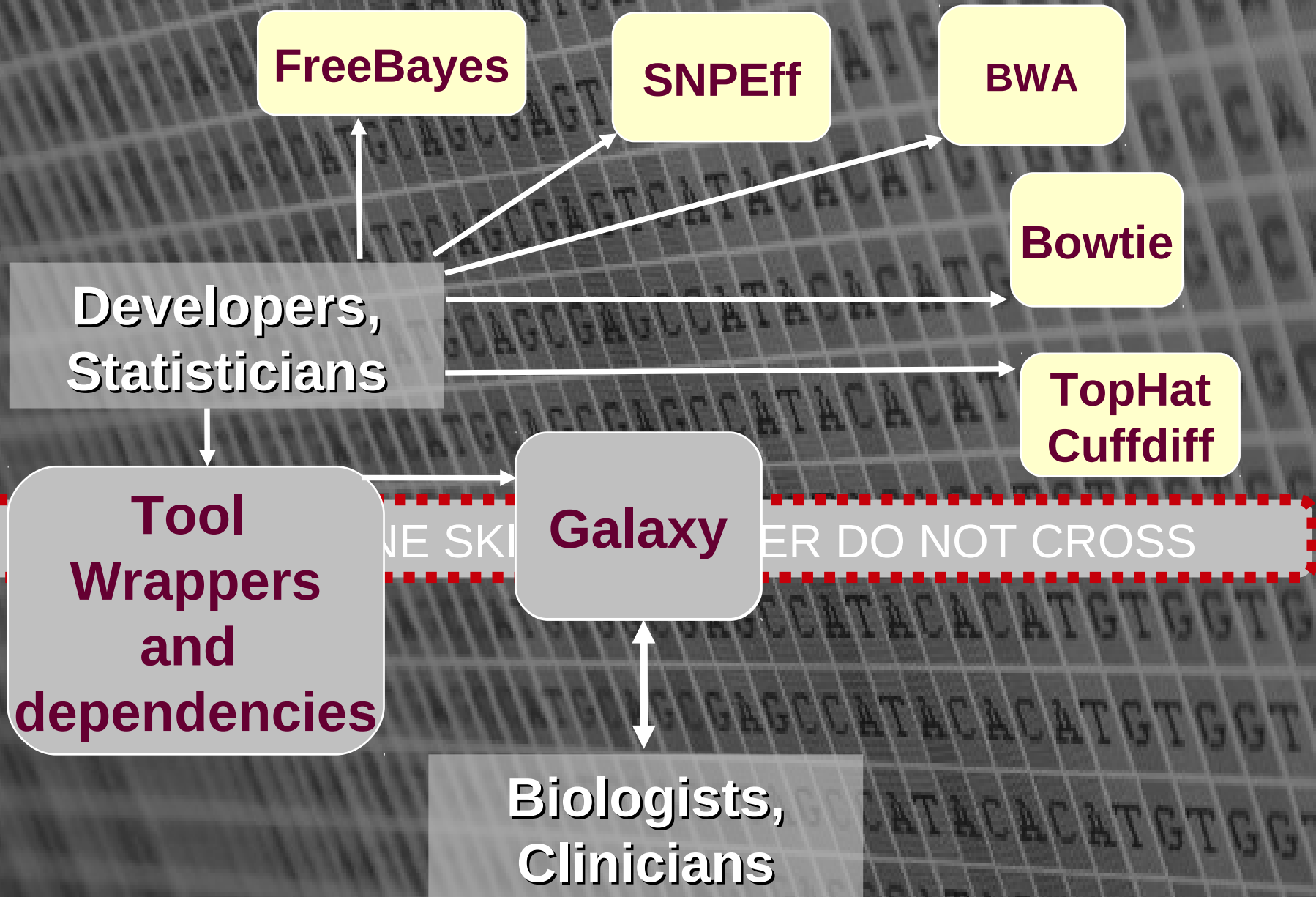
New tool distribution model

- Tool sheds ← tools and dependencies
- Galaxy distribution – few or no tools
- Galaxy admin: sees valid tool shed tools
- One click install → with all dependencies

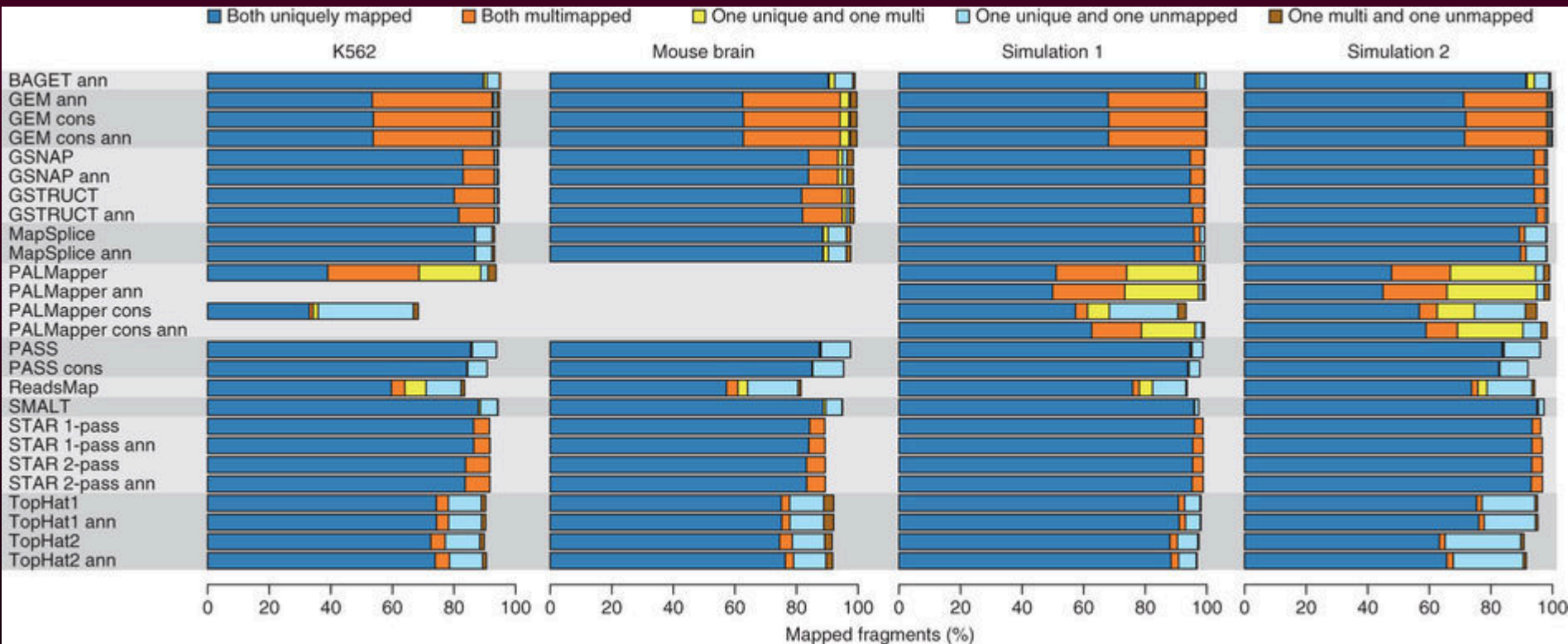
What are dependencies?

- OS projects: Tophat, Bowtie, bwa,...,etc
- Tool devs write wrapper → UI for tool
- Rapid development; regular updates
- eg bwa 0.5.9rc16 is updated to 0.6.2
- If dependency updated, different results
- Reproducibility ← dependency control

Dependencies – regular updates



Competition within task areas!



http://www.nature.com/nmeth/journal/v10/n12/fig_tab/nmeth.2722_F1.html

Old style dependencies

- Wrapper version # controlled in XML
- Tool executable was system version
- eg bwa 0.5.9
- If updated, tool calls new version silently
- System updates break reproducibility
- Researchers want latest versions
- Sysadmin under pressure...

New Tool Shed dependencies

- Tool devs create dependency packages
- Update as become available
- Tool devs write XML wrapper → UI for tool
- All dependencies specified in XML
- Strict tool/dependency version control
- Old jobs rerun with old versions
- New jobs run with latest version

Tool shed tools

- Wrapper plus specific dependencies
- Packaged as a tgz/zip/bz2 archive
- Uploaded to a new Tool Shed repository
- Changed wrapper version?
- New tool version available in Tool Shed
- Admin sees when updates available
- Click to install or update or upgrade

Tool Shed

- Independent server in Galaxy source
- Team runs main and test Tool Sheds
- Anyone can run a tool shed!
- Developers can upload to tool sheds
- Users can review and rate tools
- Every version maintained in a VCS

Demonstration Plan

- Silly example – but could be any code!
- Will use laptop local servers
- Install a tool from a toolshed
- Run the new tool
- Explore archive structure and code
- Show how tool was generated
- If time, show tool upgrade process

Outline

- Galaxy tools
- Tool Shed: VCS for tools
- Strict dependency control
- Low impedance distribution
- Demonstration
- Commodity reproducible analysis

Tool Shed: strict dep control

- Tool Shed is the app store for Tools
- Uses the mercurial VCS for storage
- Provides/accepts zip/tgz/gz archives
- Specific name/layout requirements
- Deps in tool_configuration.xml
- Galaxy maintains isolated versions
- Old jobs use same (old) version

Tool Shed

- App store for Galaxy tools
- Admin can click to install
- Explicit tool versioning
- Enhanced reproducibility
- Enhanced sharing of tools



VIVA LA EVOLUCIÓN

GALAXY

<http://usegalaxy.org>