



refinery-platform



@geoffmnelson



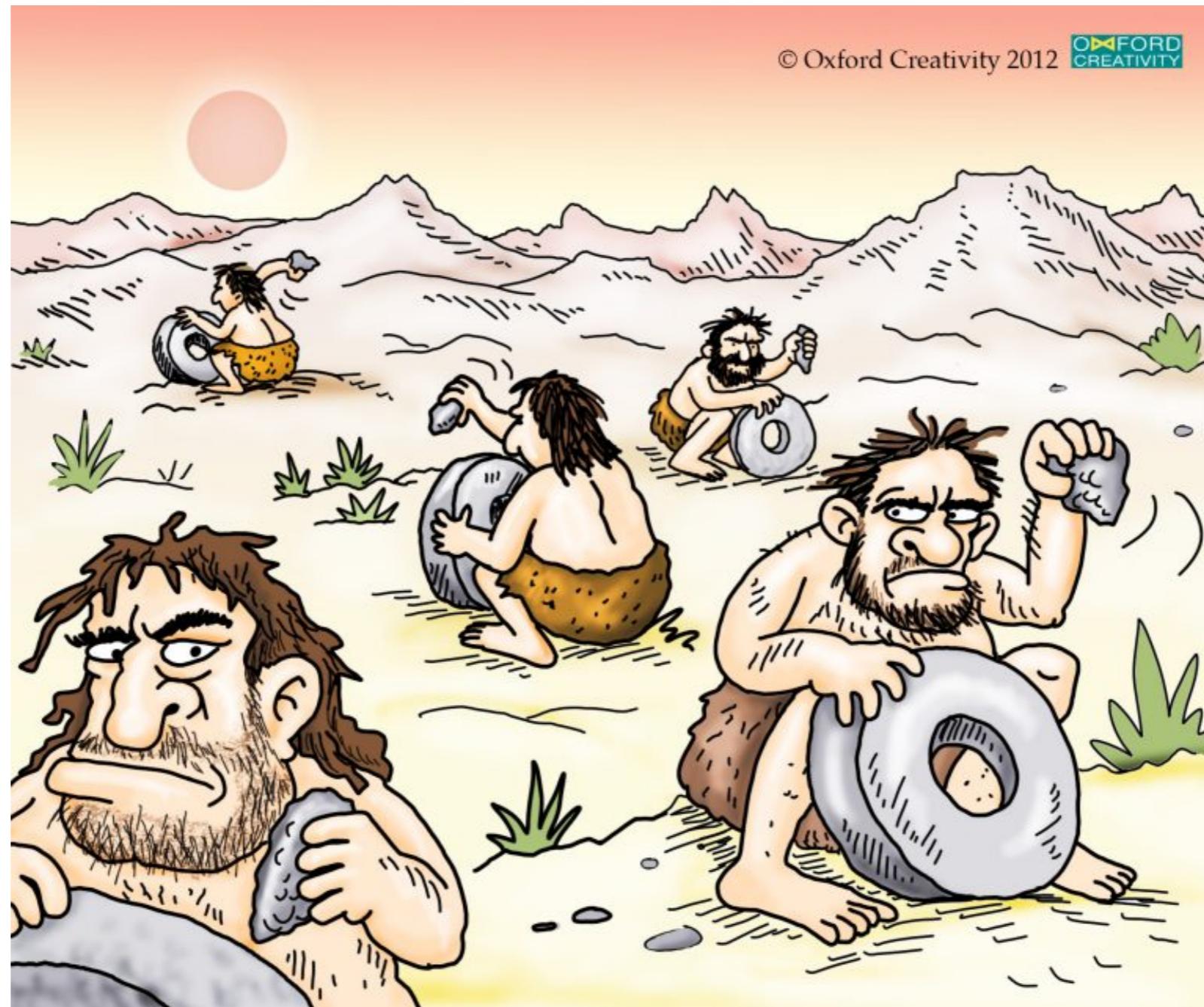
A DATA MANAGEMENT, ANALYSIS, AND
VISUALIZATION PLATFORM USING THE GALAXY
WORKBENCH

REFINERY

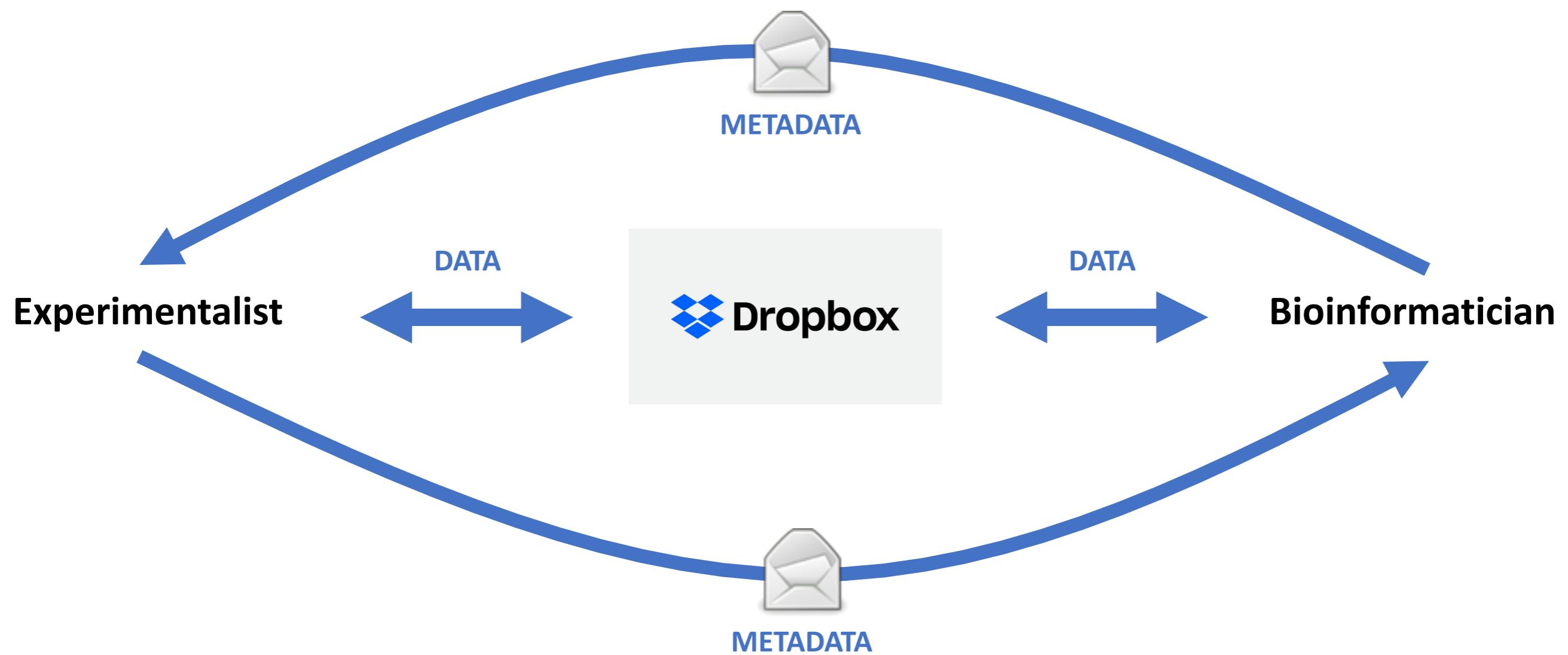
PARK LAB | COMPUTATIONAL GENOMICS



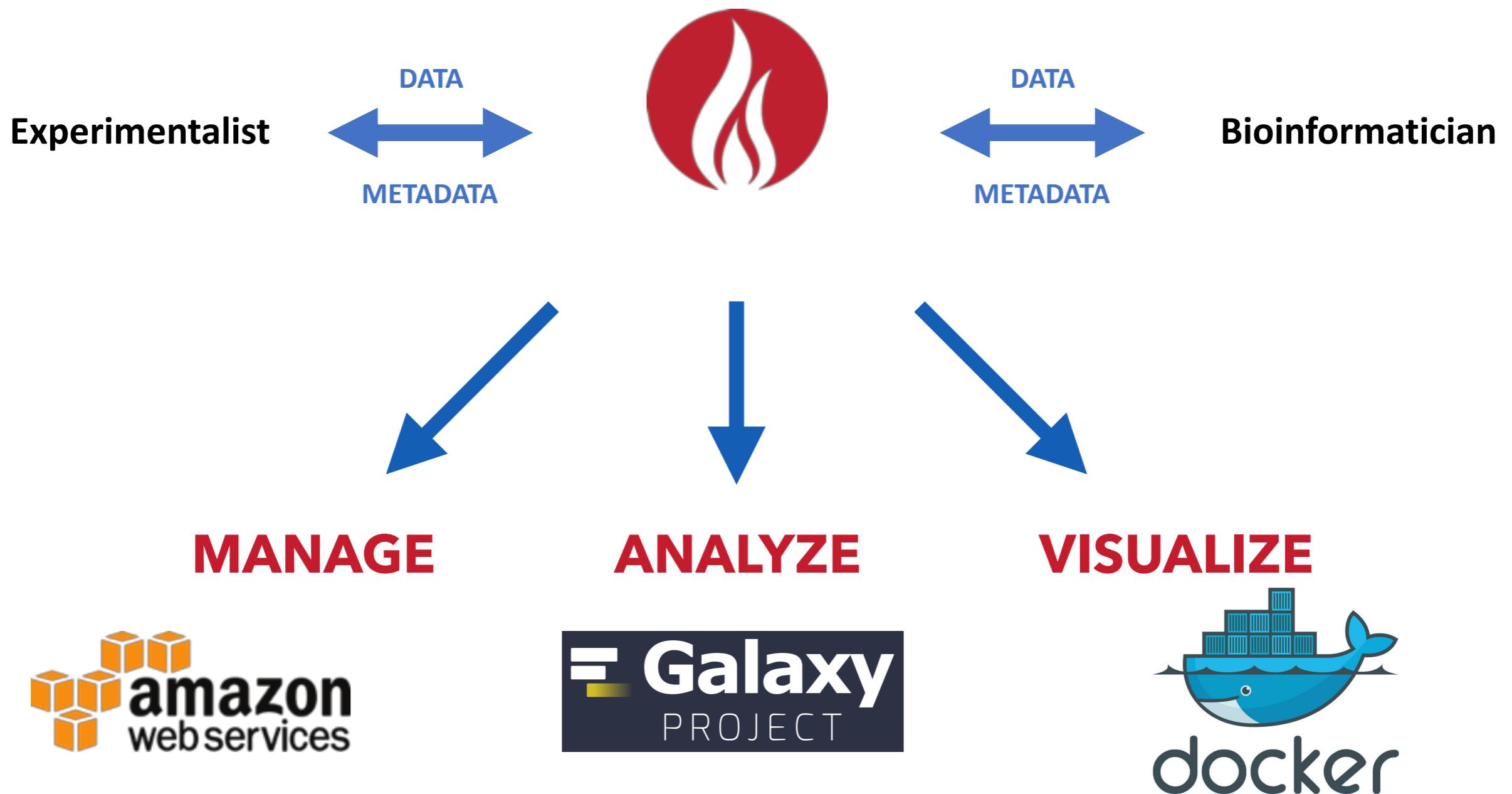
MOTIVATION



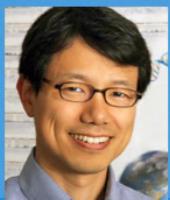
MOTIVATION



OVERVIEW



TEAM



Peter Park

Principal Investigator
Park Lab



Shannan Ho Sui

Co-investigator
Harvard Chan Bioinformatics Core



Nils Gehlenborg

Co-investigator
Gehlenborg Lab



Ilya Sytchev

Senior Bioinformatics Developer
Harvard Chan Bioinformatics Core



Jennifer Marx

Software Engineer
Gehlenborg Lab and Park Lab



Scott Ouellette

Software Developer
Gehlenborg Lab and Park Lab



Chuck McCallum

Software Developer
Gehlenborg Lab



Fritz Lekschas

Graduate student
Gehlenborg Lab



Geoff Nelson

HSCI Liaison
Park Lab

MANAGE

DASHBOARD

Stem Cell Commons Collaboration Statistics About CSCB Home

My Dashboard | Help | Geoff Nelson | Logout | ☰

Data Sets

Primary Group: Testing ? Owned Groups ✓ Public List Upload

Search

127 data sets

Signaling pathways driving leukemic stem cell fitness

We hypothesize that if secondary mutations that are advantageous to growth occur, these cells will become overrepresented both within the labeled population and the leukemic population in general. Since labeling malignant cells using fluorescent proteins allows the tracking of multiple live cell populations by flow cytometry, this system allows us to "capture" clones with a growth advantage for further characterization. In a preliminary experiment, we derived fluorescently labeled clonal cell lines from a MLL-AF9 transduced myeloid progenitor and injected them in competition in multiple murine replicates. By tracking the growth of leukemic cells in the peripheral blood, which we found to be representative of total leukemic burden, we observed a stochastic pattern of clonal evolution. Isolation of live leukemic stem cell-enriched populations within each label, for DNA, RNA and ex vivo culture, is possible throughout disease evolution by FACS-sorting femoral bone marrow aspirates obtained from these mice at monthly intervals. Our preliminary data demonstrate that "winning" clones have an increased clonogenic potential in colony-forming assay. Data originally uploaded by Jiantao Shi.

May 18, 2018

Osetoprogenitor Dicer KO LKS cells

Data set originally uploaded by Youmna Kfouri.

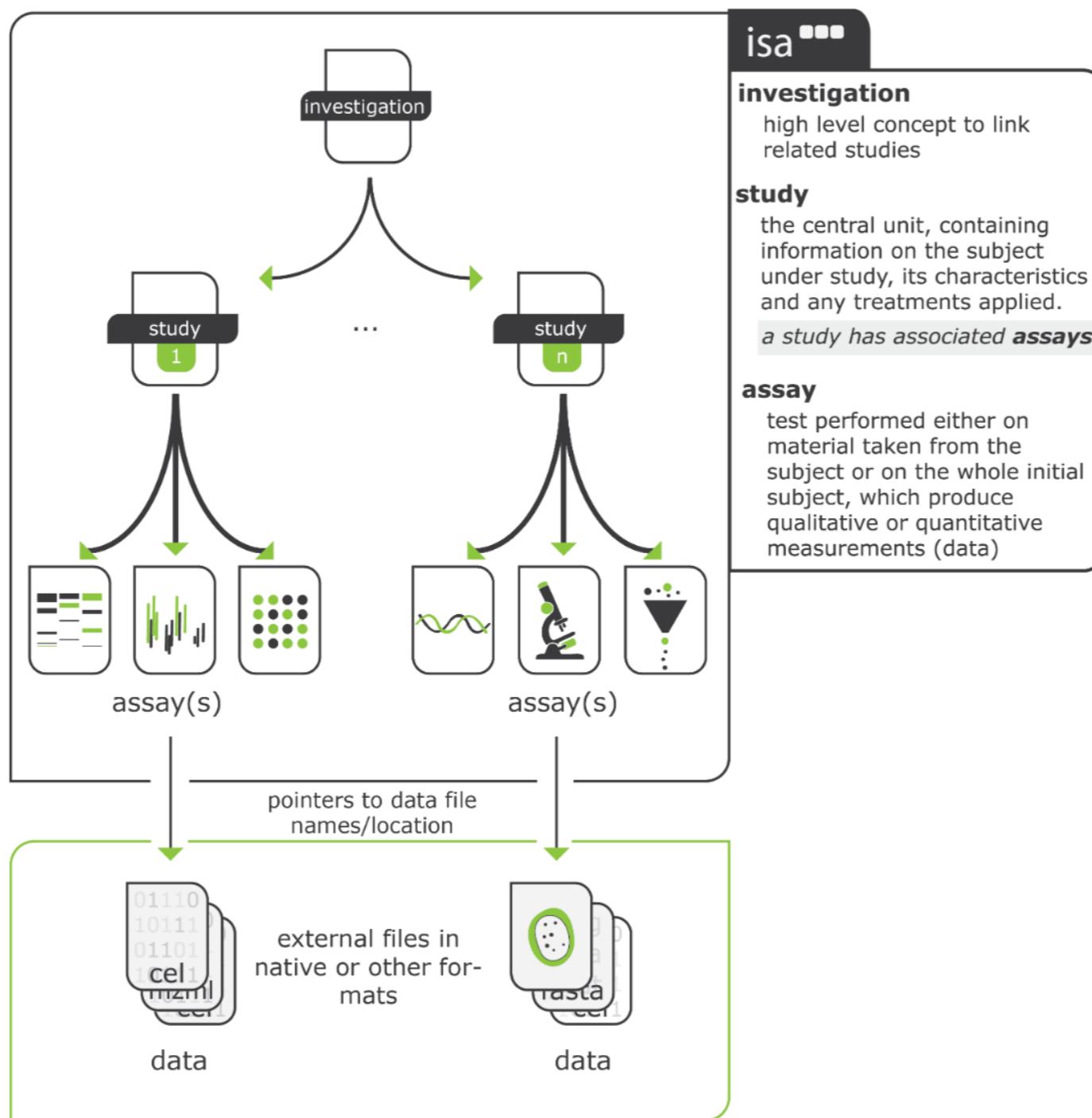
History

- 07/23/2018 Geoff Nelson launched an analysis ChIP-seq Peak Calling - Mouse 2018-07-23 14:17:43.261199+00:00 gnelson on data set H3K79me2 ChIP-seq in mouse proximal intestinal Lgr5(hi) stem cells and villus cells
- 07/23/2018 Geoff Nelson created data set H3K79me2 ChIP-seq in mouse proximal intestinal Lgr5(hi) stem cells and villus cells
- 07/21/2018 Scott Ouellette launched an analysis FastQC 2018-07-21 04:34:14.054812+00:00 scott_ouellette on data set tutorial
- 07/21/2018 Scott Ouellette launched visualization MultiQC 2018-07-21 04:31:59.521354+00:00 scott_ouellette on data set tutorial

Collaboration

Group	Role	Edit
Testing		
Shivdasani Lab	Manager	
Breault Lab	Manager	
Public		
Tutorial	Manager	
Rosen Lab	Manager	

METADATA | ISA-TAB



isa-tools.org



ISA-tools

PROVENANCE

Stem Cell Commons Collaboration Statistics About CSCB Home My Dashboard | Help | Geoff Test | Logout |

Data Set rna-seq_wt_vs_ko: Transcriptional profiles of WT and KO mice

Files Analyses Visualizations Details Provenance

Layers Analyses Analysis Groups Workflows Attributes Filter Blend Hide Info panel Scale graph ?

```
graph LR; subgraph KO2 [KO2]; D1[dataset KO2] --> F1[KO2.R1.fastq.gz]; F1 --> S1[salmon4]; S1 --> Q1[KO2.quant.sf]; end; subgraph KO1 [KO1]; D2[dataset KO1] --> F2[KO1.R1.fastq.gz]; F2 --> S2[salmon3]; S2 --> Q2[KO1.quant.sf]; end; subgraph WT2 [WT2]; D3[dataset WT2] --> F3[WT2.R1.fastq.gz]; F3 --> S3[salmon2]; S3 --> Q3[WT2.quant.sf]; end; subgraph WT1 [WT1]; D4[dataset WT1] --> F4[WT1.R1.fastq.gz]; F4 --> S4[salmon1]; S4 --> Q4[WT1.quant.sf]; end; F1 --> M2((merge-counts2)); F2 --> M2; F3 --> M1((merge-counts1)); F4 --> M1; M2 --> K[KO-counts.tsv]; M1 --> W[WT-counts.tsv]; K --> D[deseq2]; W --> D; D --> O[WTVsko-de.tsv]
```

ANALYZE

WORKFLOW LAUNCH

Stem Cell Commons [Collaboration](#) [Statistics](#) [About](#) [CSCB Home](#)

[My Dashboard](#) | [Help](#) | [Geoff Nelson](#) | [Logout](#) | [⚙️](#)

Data Set 16182: H3K79me2 ChIP-seq in mouse proximal intestinal Lgr5(hi) stem cells and villus cells

« Show Tool Panel [⟳](#) [🔗](#)

7 FILES [Revise Metadata](#) [?](#)

File	Analysis Group	Group Name	Strain	Antibody	Replicate Id	Organism	Name
⬇️	N/A	villus H3K79me2_ChIP...	C57BL/6	H3K79me2	15468	Mus musculus	ftp://ftp.sra.ebi.a
⬇️	N/A	viillus input DNA	C57BL/6	none	15467	Mus musculus	ftp://ftp.sra.ebi.a
⬇️	N/A	Lgr5(hi) H3K79me2_Ch...	C57BL/6	H3K79me2	15470	Mus musculus	ftp://ftp.sra.ebi.a
⬇️	N/A	Lgr5(hi) input DNA	C57BL/6	none	15469	Mus musculus	ftp://ftp.sra.ebi.a
⬇️	0	Lgr5(hi) H3K79me2_Ch...	C57BL/6	H3K79me2 + none	15467 + 15468 + 15469 + ...	Mus musculus	MACS2 callpeak .
⬇️	0	Lgr5(hi) H3K79me2_Ch...	C57BL/6	H3K79me2 + none	15467 + 15468 + 15469 + ...	Mus musculus	MACS2 bdgcmp .
⬇️	0	Lgr5(hi) H3K79me2_Ch...	C57BL/6	H3K79me2 + none	15467 + 15468 + 15469 + ...	Mus musculus	Wig/BedGraph-t.

ATTRIBUTE FILTER

- search...
- ▶ Analysis Group (2)
- ▶ Antibody (3)
- ▶ Output Type (4)
- ▶ Type (2)
- ▶ Cell Type (3)
- ▶ FileType (4)

ANALYSIS FILTER

- ▶ Analysis

IMPLEMENTATION

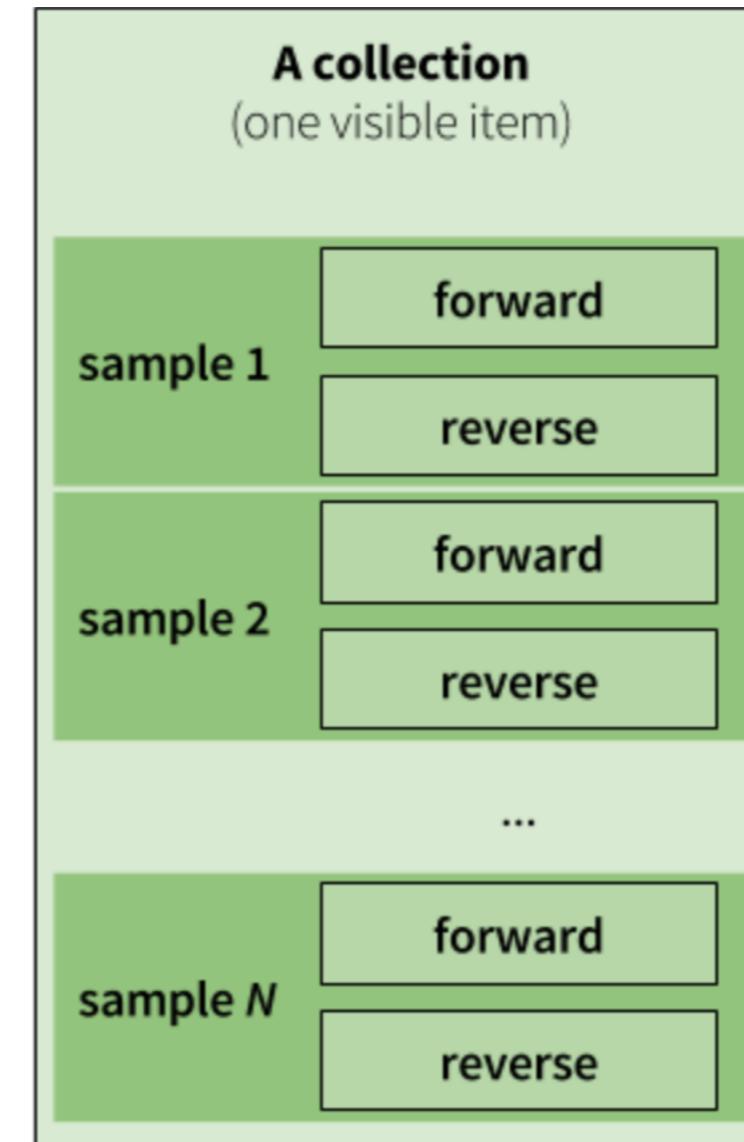
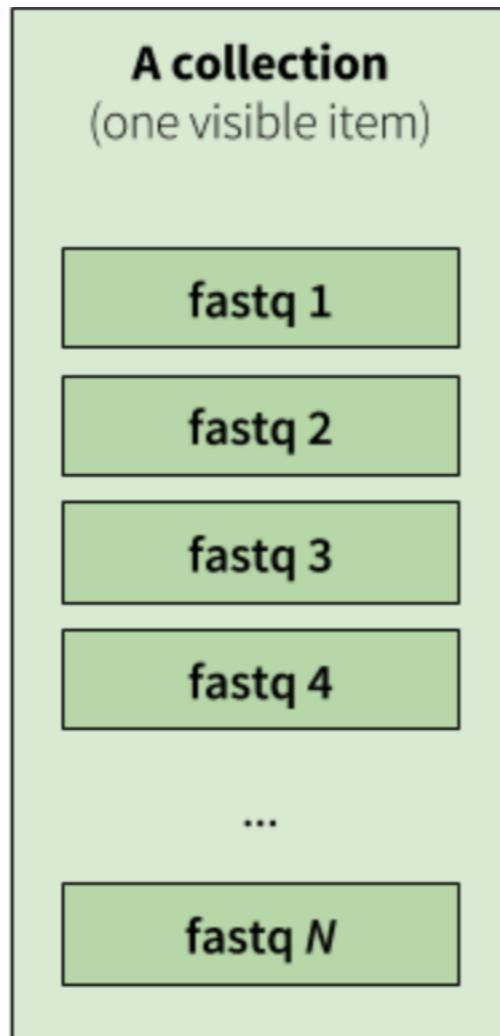


BioBlend



[stemcellcommons/scc-galaxy-tools](https://github.com/stemcellcommons/scc-galaxy-tools)

COLLECTIONS

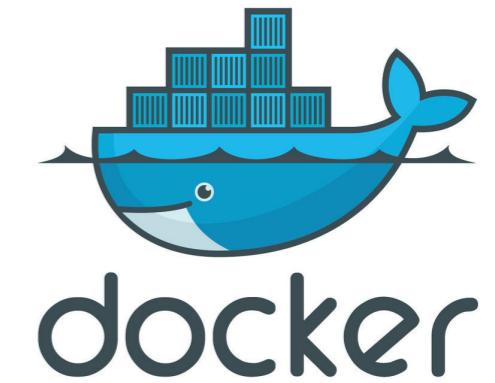


AVAILABLE WORKFLOWS

- ▶ **QC**
 - ▶ FastQC
 - ▶ Qualimap + deepTools BAM QC
- ▶ **RNA-seq**
 - ▶ Kallisto
- ▶ **ChIP-seq**
 - ▶ MACS2

VISUALIZE

IMPLEMENTATION

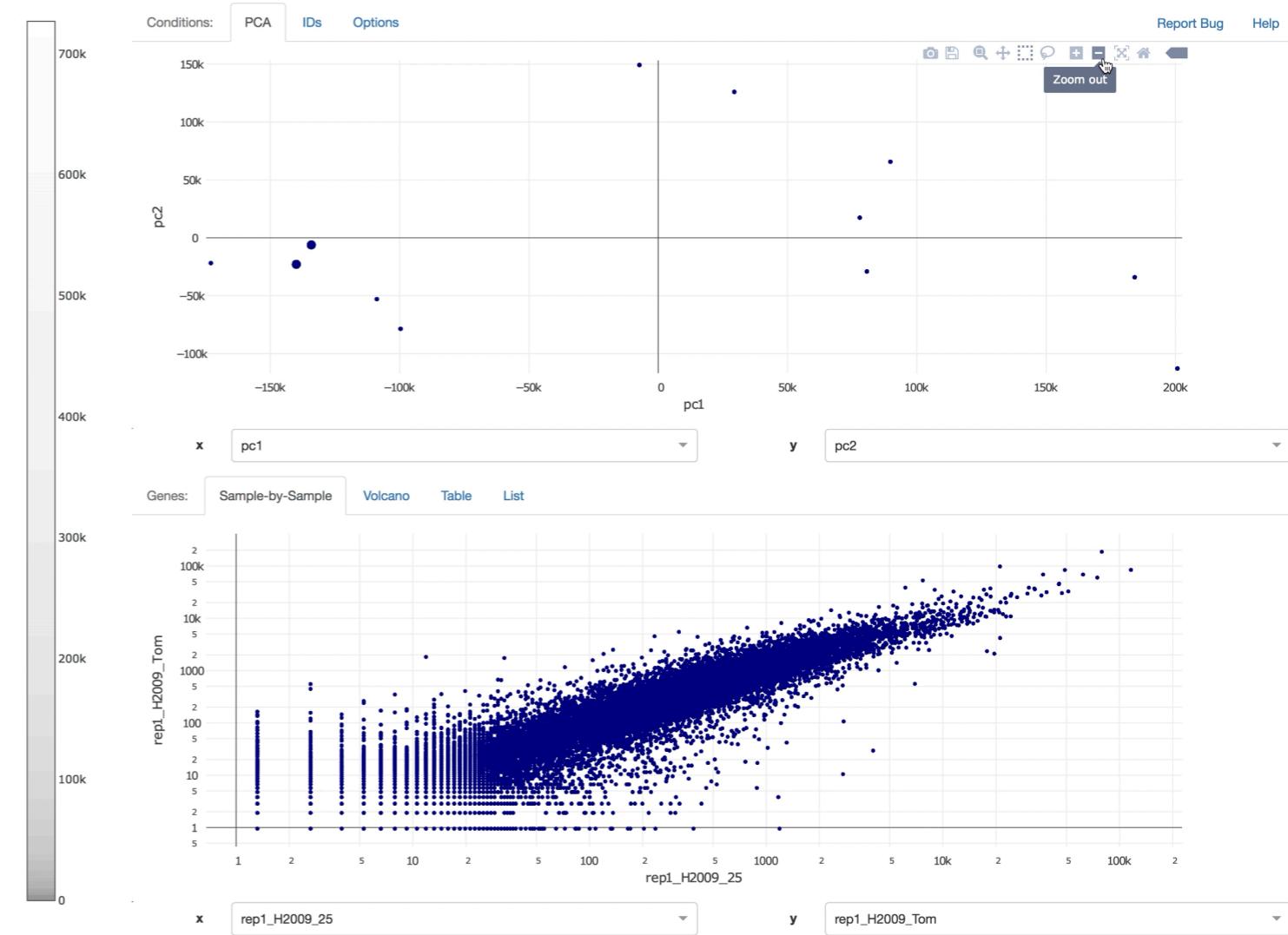
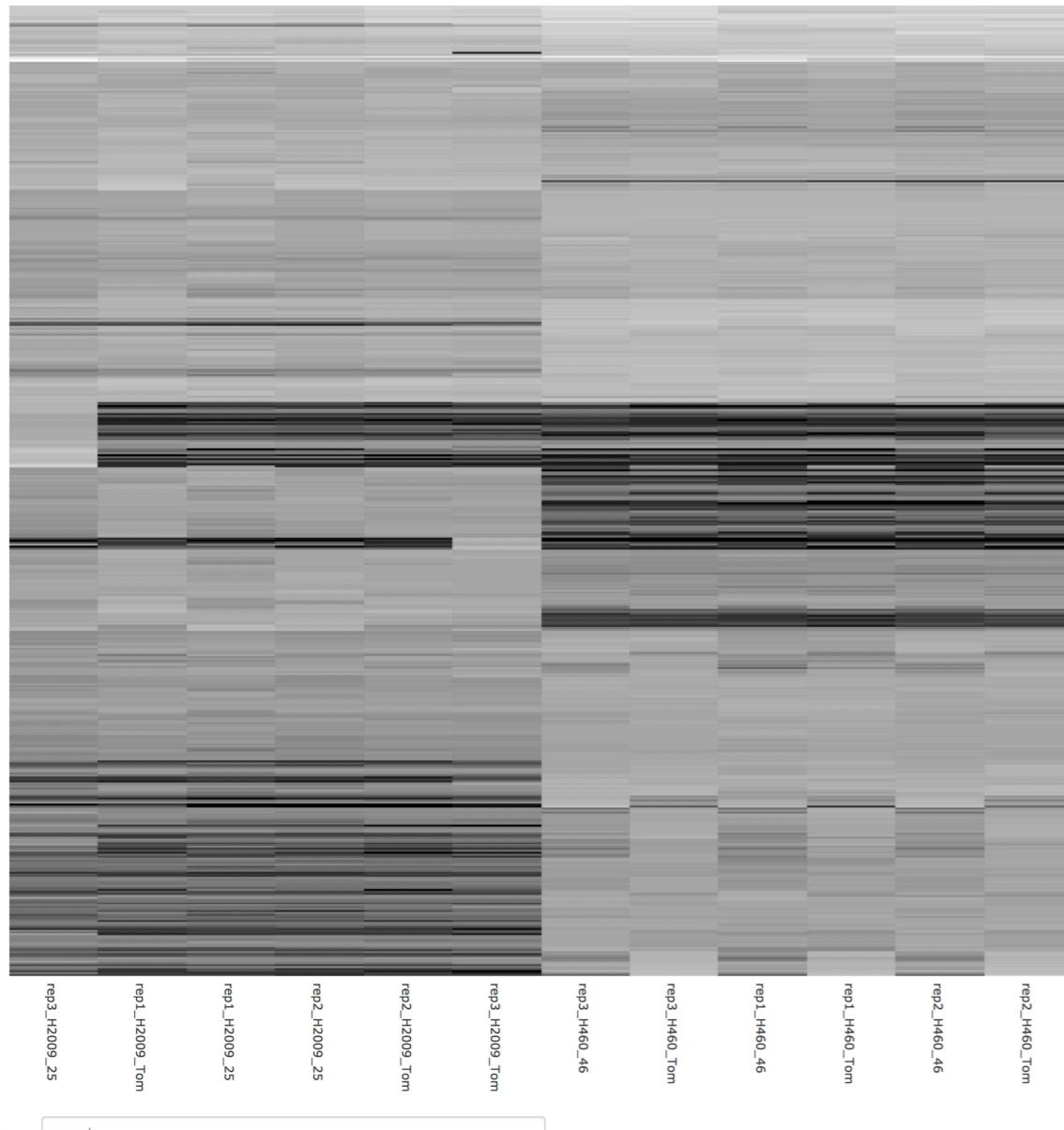


`django_docker_engine`



[refinery-platform/django_docker_engine](https://github.com/refinery-platform/django_docker_engine)

RNA-SEQ VISUALIZATION



INTERVENE | SET INTERSECTIONS

Welcome to Intervene's Shiny App!

Intervene Shiny App provides an interactive interface for intersection and effective visualization of gene or genomic region sets. Currently, Shiny app does not accept genomic regions as input, but the text files generated by Intervene's command line interface can be easily uploaded to further explore and customize the plots in an interactive way. Intervene has three modules: `venn` to generate Venn diagrams of up-to 6 sets, `upSet` to generate UpSet plots of more than 3 sets and `pairwise` to compute and visualize pairwise intersections as clustered heatmap.

Venn module

Intervene's venn module provides up-to 6-way classical, Chow-Ruskey and Edwards' Euler/Venn diagrams to visualize the intersections of genomic regions or lists.

UpSet module

Intervene's UpSet modules can be used to visualize the intersection of multiple genomic region sets using UpSet plots.

Intersection	Size
H3K27me3	17505
H3K4me3	15678
H3K27ac	15021
H3K4me2	12011
H3K27me3 & H3K4me3	7235
H3K27me3 & H3K27ac	6777
H3K27me3 & H3K4me2	5909
H3K4me3 & H3K27ac	3852
H3K4me3 & H3K4me2	2216
H3K27ac & H3K4me2	1698
H3K27me3 & H3K27ac	1490
H3K27me3 & H3K4me2	473
H3K4me3 & H3K4me2	307
H3K27ac & H3K4me2	295
H3K27me3 & H3K4me3 & H3K27ac	256
H3K27me3 & H3K4me3 & H3K4me2	0
H3K27me3 & H3K27ac & H3K4me2	0
H3K4me3 & H3K27ac & H3K4me2	0
H3K27me3 & H3K4me2 & H3K4me2	0

Pairwise module

Intervene's pairwise module provides several styles of heatmaps and clustering approaches to customize the heatmaps.

Citation

If you use intervene, please cite this paper:

Khan A, Mathelier A. *Intervene: a tool for intersection and visualization of multiple gene or genomic region sets*. BMC Bioinformatics. 2017;18:287. doi: 10.1186/s12859-017-1708-7

LINEUP | TABLE EXPLORATION

This visualization is a Table Exploration tool, likely from Tableau, showing player statistics across various leagues and clubs. The main area displays a table of player data with numerous filters and summaries on the right side.

Table Headers:

- Row Number (Ra...)
- Player
- # Age
- Current_club
- Current_league
- Foot
- # Height
- Nationality
- Position
- Games
- Goals
- Minutes
- Assists

Key Observations:

- The table contains 1177 rows of data, with the first few rows including Dmitri Poloz, Emiliano Rigoni, and Sebastián Driussi.
- Players are primarily from Russia, Argentina, Italy, and Brazil.
- Positions include CF (Center Forward), W (Winger), and various defensive roles like CB (Center Back) and RB (Right Back).
- Clubs listed include Zenit St. Petersburg, Spartak Moscow, CSKA Moscow, FK Krasnodar, Lokomotiv Moscow, Rubin Kazan, Akhmat Grozny, Dinamo Moscow, and Arsenal Tula.
- Leagues represented include Premier League (UK), Eredivisie (Netherlands), Jupiler Pro League (Belgium), La Liga (Spain), Liga MX (Mexico), Premer Lig (Turkey), Serie A (Italy), Supercoppa Italiana (Italy), and Süper Lig (Turkey).
- Foot types shown are both, left, and right.
- Height ranges from approximately 179 cm to 204 cm.
- Age ranges from 22 to over 30 years old.

Right Panel (Filters and Summaries):

- Item Overview: Shows 1177 of 1177 items.
- Add Column...
- Column summaries
- Filter Player...
- RegExp
- Filter rows containing missing values
- # Age: Histogram showing age distribution.
- Current_club: Bar chart showing the count of players by club.
- Filter Current_club...
- RegExp
- Filter rows containing missing values
- Current_league: Bar chart showing the count of players by league.
- Filter 0 remaining missing value rows
- Foot: Bar chart showing the count of players by foot type.
- Filter 53 remaining missing value rows
- # Height: Histogram showing height distribution.
- Filter 0 remaining missing value rows
- Nationality: Bar chart showing the count of players by nationality.
- Filter 0 remaining missing value rows



refinery-platform

HSCI HARVARD
STEM CELL INSTITUTE

1000 scientists. One goal. Discovering cures.

