REFINERY PLATFORM

Integrating Visualization *and* Analysis *of* Large-Scale Biological Data

NILS GEHLENBORG

Harvard Medical School
Center for Biomedical Informatics

Broad Institute of MIT and Harvard
Cancer Program

REFINERY PLATFORM

PARK LAB
Harvard Medical School
Center for Biomedical Informatics

Nils Gehlenborg
Richard Park
Psalm Haseley
Peter J Park

HIDE LAB
Harvard School of Public Health
Department of Biostatistics

Shannan Ho Sui
Ilya Sytchev
Winston Hide

Funded in part by
Agilent Technologies

TCGA

The Cancer Genome Atlas

20 cancer types

×

500 patients

10,000 patients

mRNA expression

microRNA expression

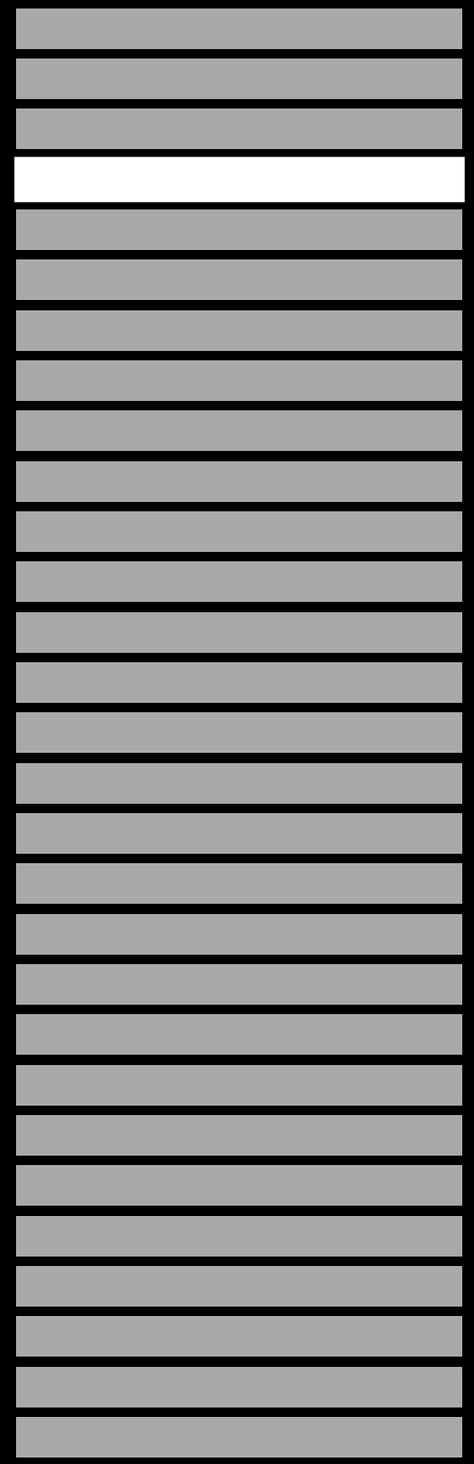
DNA methylation

protein expression

copy number variants

mutation calls

clinical parameters

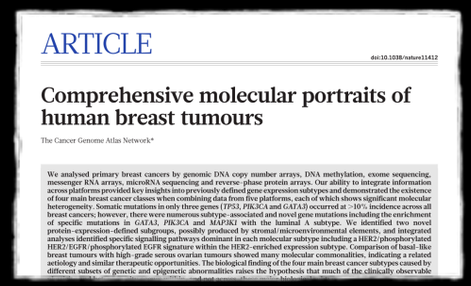


mRNA expression

microRNA expression

DNA methylation

protein expression



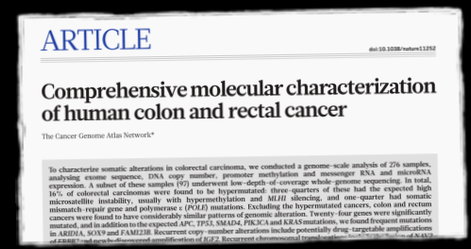
ARTICLE

doi:10.1038/nature11412

Comprehensive molecular portraits of human breast tumours

The Cancer Genome Atlas Network*

We analysed primary breast cancers by genomic DNA copy number arrays, DNA methylation, exome sequencing, messenger RNA arrays, microRNA sequencing and reverse-phase protein arrays. Our ability to integrate information across platforms provided key insights into previously defined gene expression subtypes and demonstrated the existence of four main breast cancer classes when combining data from five platforms, each of which shows significant molecular heterogeneity. Somatic mutations in only three genes (*TP53*, *PIK3CA* and *GATA3*) occurred at >10% incidence across all breast cancers; however, there were numerous subtype-associated and novel gene mutations including the enrichment of specific mutations in *GATA3*, *PIK3CA* and *MAP3K1* with the luminal A subtype. We identified two novel protein-expression-defined subgroups, possibly produced by stromal/microenvironmental elements, and integrated analyses identified specific signalling pathways dominant in each molecular subtype including a HER2/phosphorylated HER2/EGFR/phosphorylated EGFR signature within the HER2-enriched expression subtype. Comparison of basal-like breast tumours with high-grade serous ovarian tumours showed many molecular commonalities, indicating a related aetiology and similar therapeutic opportunities. The biological finding of the four main breast cancer subtypes caused by different subsets of genetic and epigenetic abnormalities raises the hypothesis that much of the clinically observable



ARTICLE

doi:10.1038/nature11252

Comprehensive molecular characterization of human colon and rectal cancer

The Cancer Genome Atlas Network*

To characterize somatic alterations in colorectal carcinoma, we conducted a genome-scale analysis of 276 samples, analysing exome sequence, DNA copy number, promoter methylation and messenger RNA and microRNA expression. A subset of these samples (97) underwent low-depth-of-coverage whole-genome sequencing. In total, 16% of colorectal carcinomas were found to be hypermutated: three-quarters of these had the expected high microsatellite instability, usually with hypermethylation and *MLH1* silencing, and one-quarter had somatic mismatch-repair gene and polymerase ε (*POLE*) mutations. Excluding the hypermutated cancers, colon and rectum cancers were found to have considerably similar patterns of genomic alteration. Twenty-four genes were significantly mutated, and in addition to the expected *APC*, *TP53*, *SMAD4*, *PIK3CA* and *KRAS* mutations, we found frequent mutations in *ARID1A*, *SOX9* and *FAM123B*. Recurrent copy-number alterations include potentially drug-targetable amplifications of *ERBB2* and newly discovered amplification of *IGF2*. Recurrent chromosomal translocations

copy number variants

mutation calls

clinical parameters

Vol 455|23 October 2008|doi:10.1038/nature07385

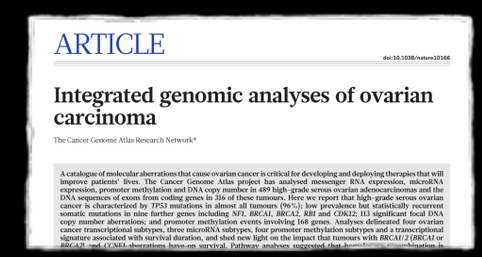
nature

ARTICLES

Comprehensive genomic characterization defines human glioblastoma genes and core pathways

The Cancer Genome Atlas Research Network*

Human cancer cells typically harbour multiple chromosomal aberrations, nucleotide substitutions and epigenetic modifications that drive malignant transformation. The Cancer Genome Atlas (TCGA) pilot project aims to assess the value of large-scale multi-dimensional analysis of these molecular characteristics in human cancer and to provide the data rapidly to the research community. Here we report the interim integrative analysis of DNA copy number, gene expression and DNA methylation aberrations in 206 glioblastomas—the most common type of primary adult brain cancer—and nucleotide sequence aberrations in 91 of the 206 glioblastomas. This analysis provides new insights into the roles of *ERBB2*, *NF1* and



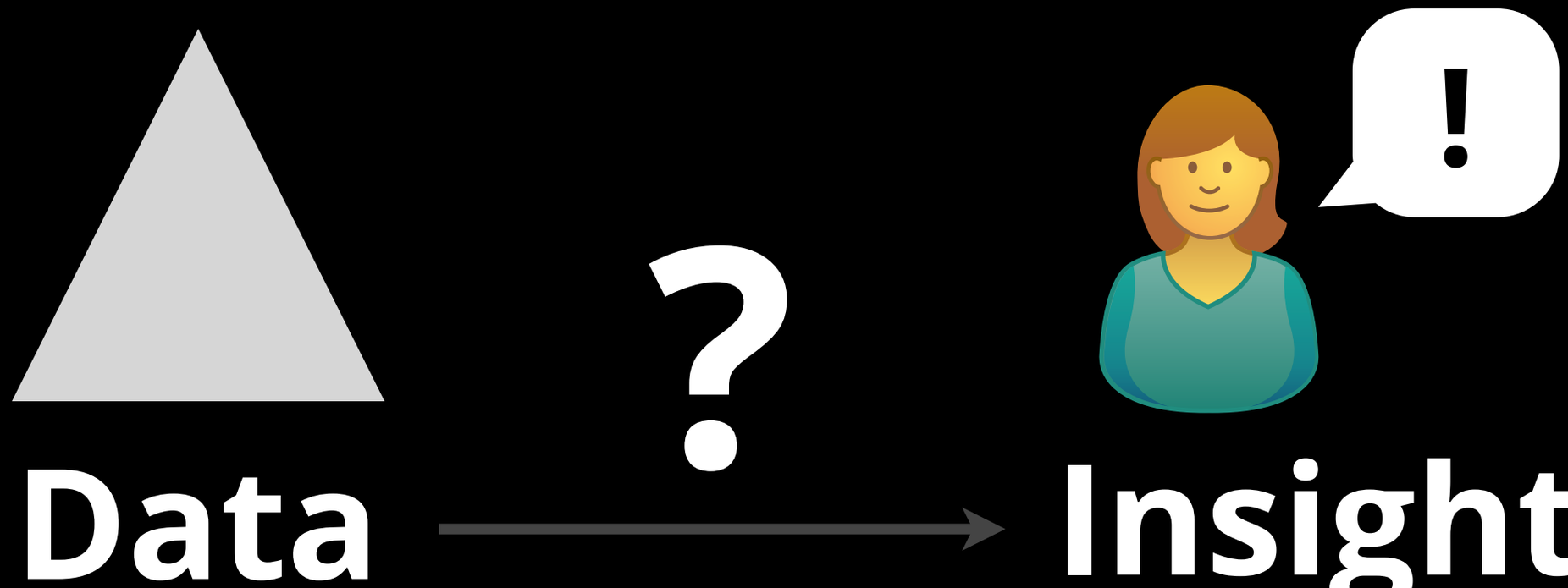
ARTICLE

doi:10.1038/nature10166

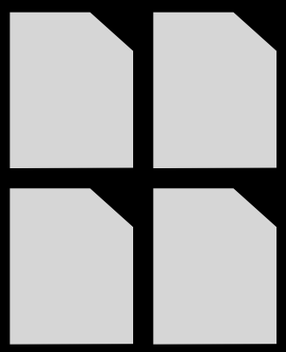
Integrated genomic analyses of ovarian carcinoma

The Cancer Genome Atlas Research Network*

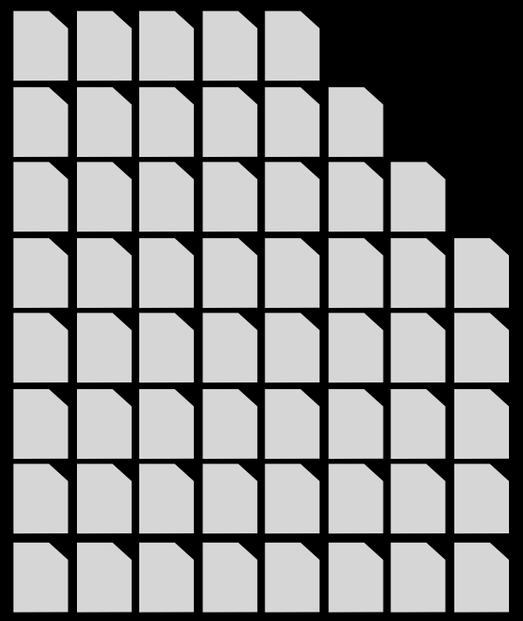
A catalogue of molecular aberrations that cause ovarian cancer is critical for developing and deploying therapies that will improve patients' lives. The Cancer Genome Atlas project has analysed messenger RNA expression, microRNA expression, promoter methylation and DNA copy number in 489 high-grade serous ovarian adenocarcinomas and the DNA sequences of exons from coding genes in 316 of these tumours. Here we report that high-grade serous ovarian cancer is characterized by *TP53* mutations in almost all tumours (96%); low prevalence but statistically recurrent somatic mutations in nine further genes including *NF1*, *BRCA1*, *BRCA2*, *RB1* and *CDK12*; 113 significant focal DNA copy number aberrations; and promoter methylation events involving 168 genes. Analyses delineated four ovarian cancer transcriptional subtypes, three microRNA subtypes, four promoter methylation subtypes and a transcriptional signature associated with survival duration, and shed new light on the impact that tumours with *BRCA1/2* (*BRCA1* or *BRCA2*) and *CCNE1* aberrations have on survival. Pathway analyses suggested that

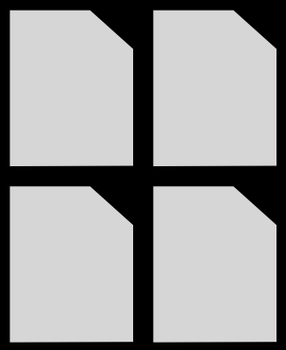

!
?
Data
Insight

Analysis Pipeline

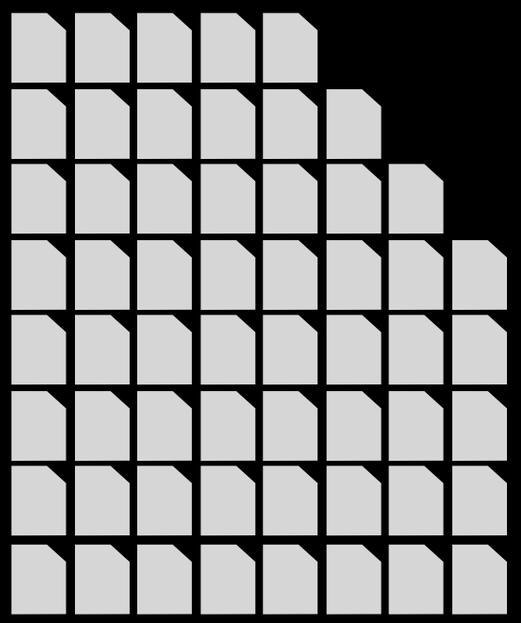


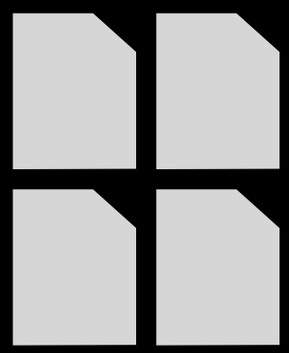
Analysis Pipeline



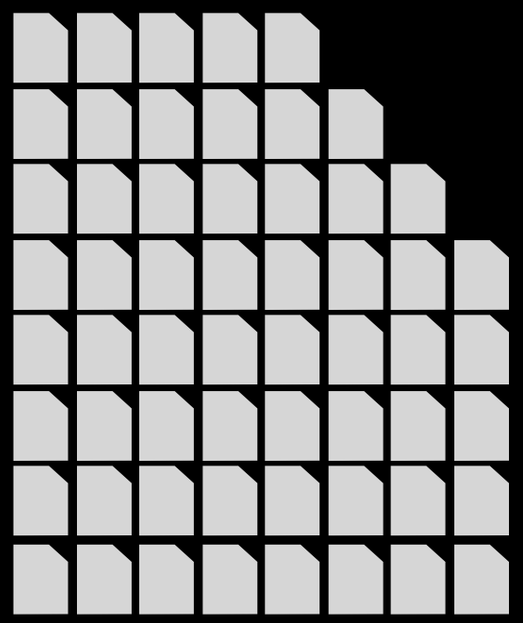


Analysis Pipeline



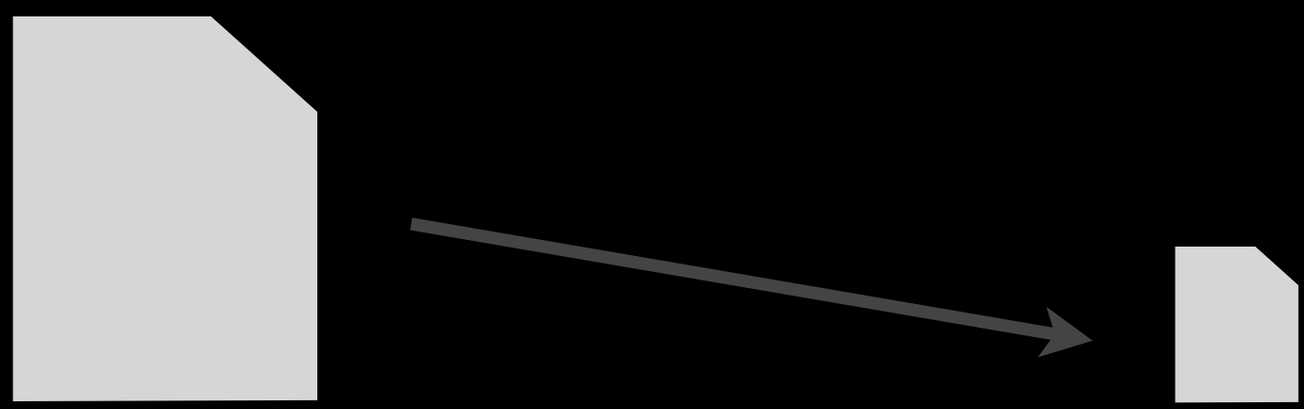


Analysis Pipeline

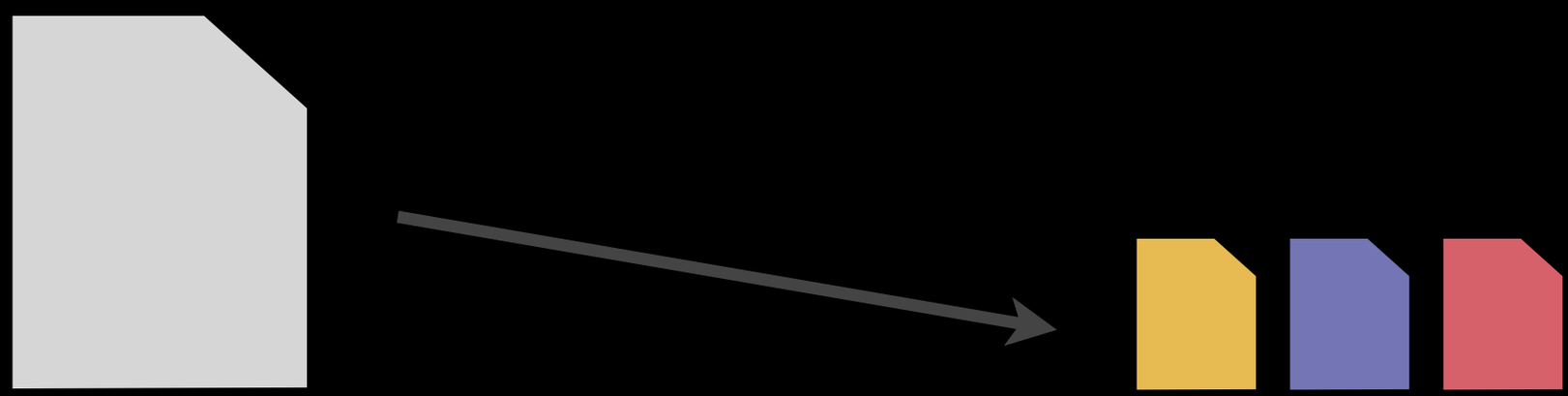


Multiple Steps

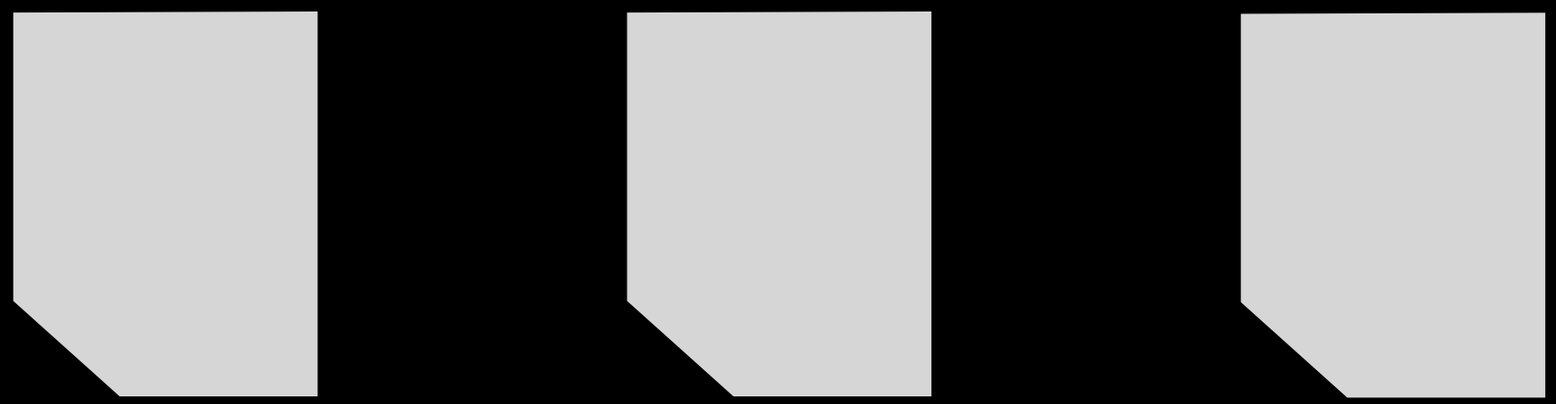
Long Runtime

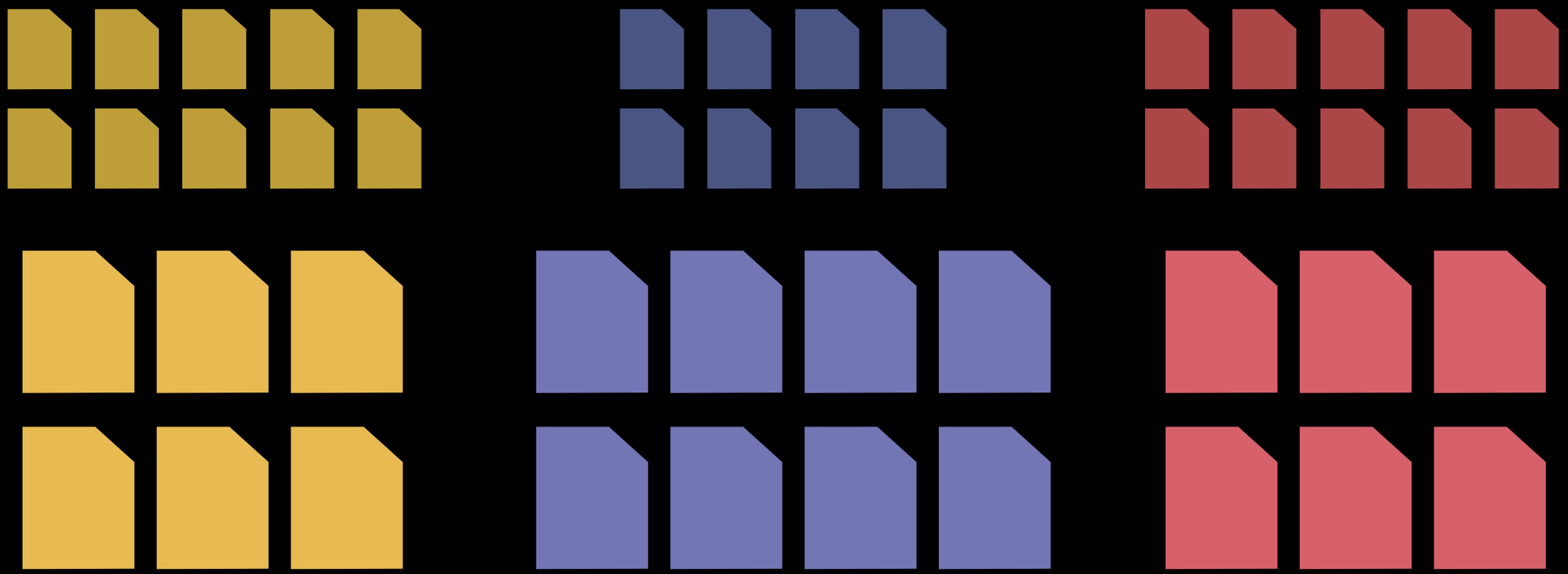


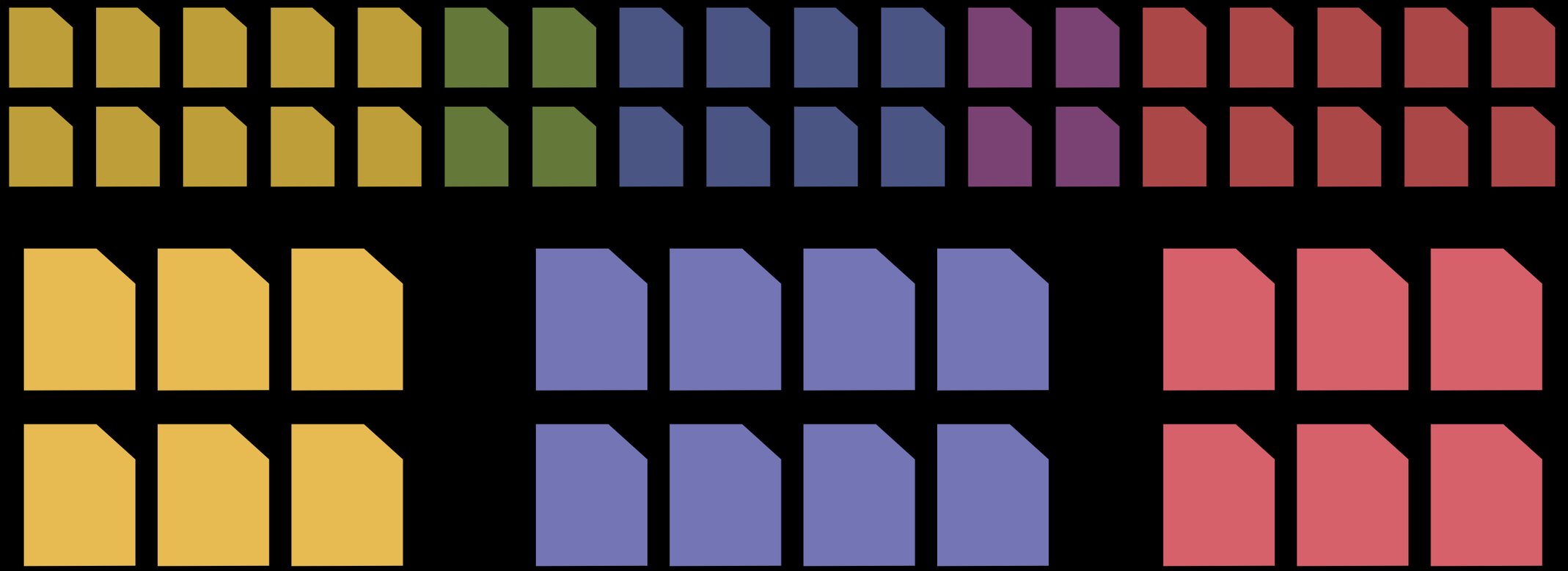
Decreasing Size



Increasing Heterogeneity







INTEGRATION

INTEGRATION

INTEGRATION

Horizontal Integration across Data Types

Biological Insight

Vertical Integration across Data Levels

Confirmation & Troubleshooting

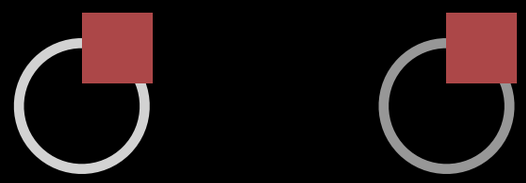
REFINERY PLATFORM

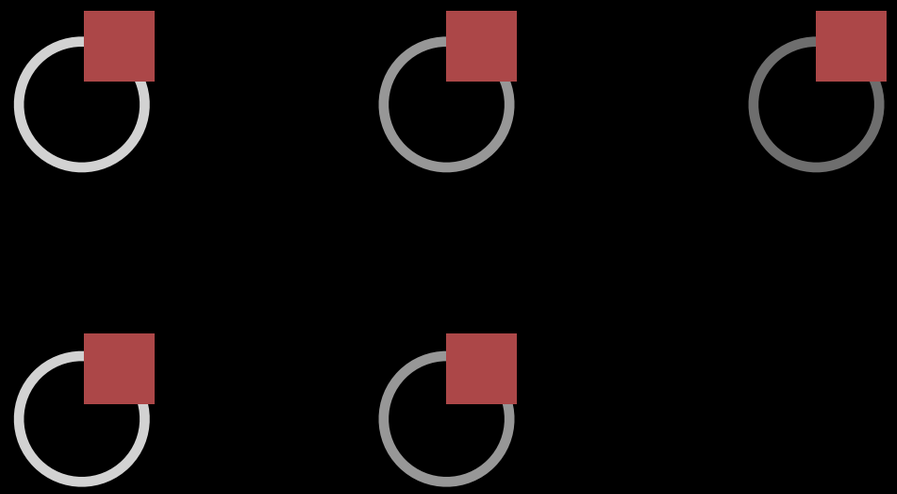
Repository

Meta data annotation

Repository

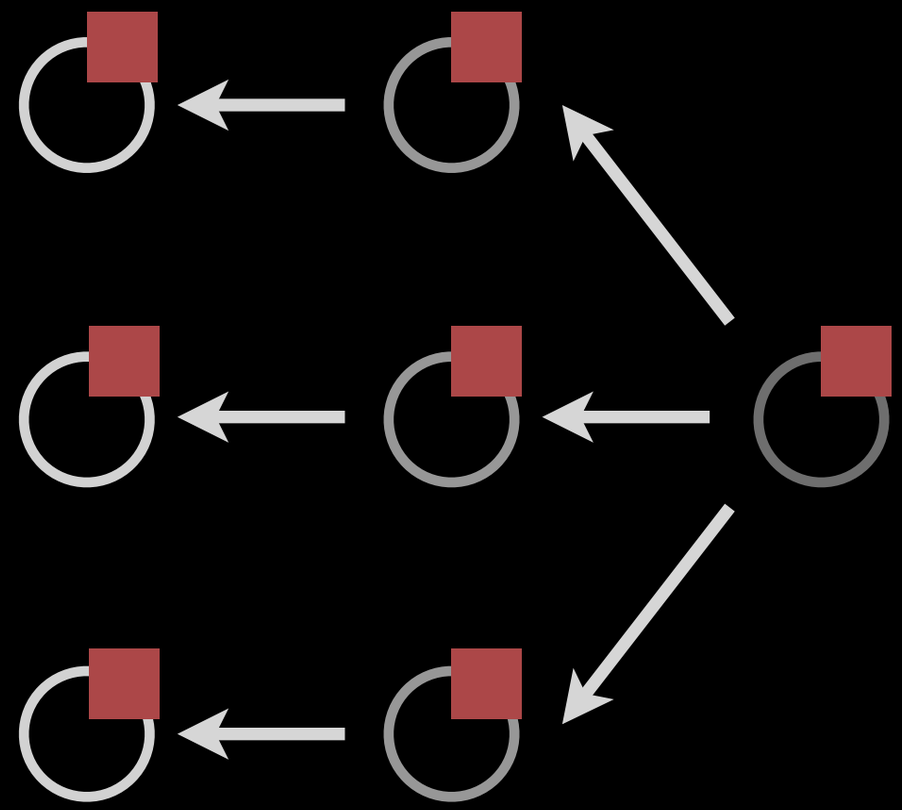
Meta data annotation





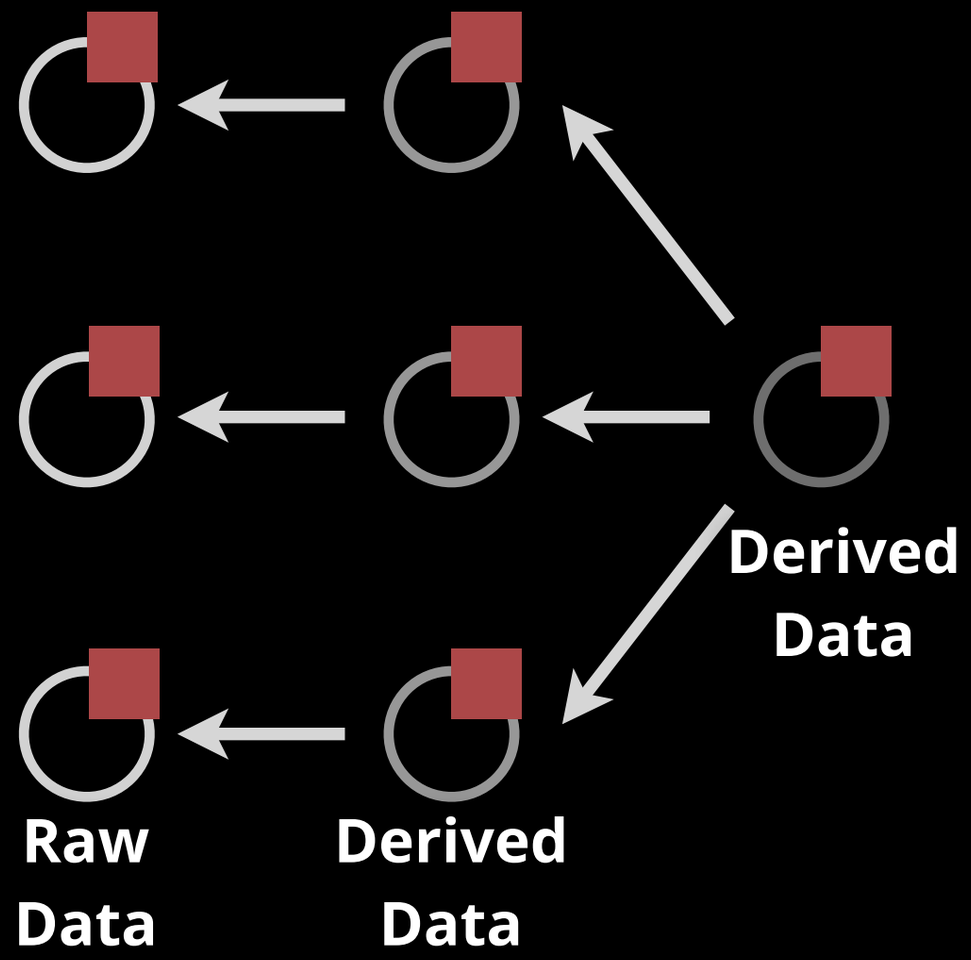
Repository

Meta data annotation



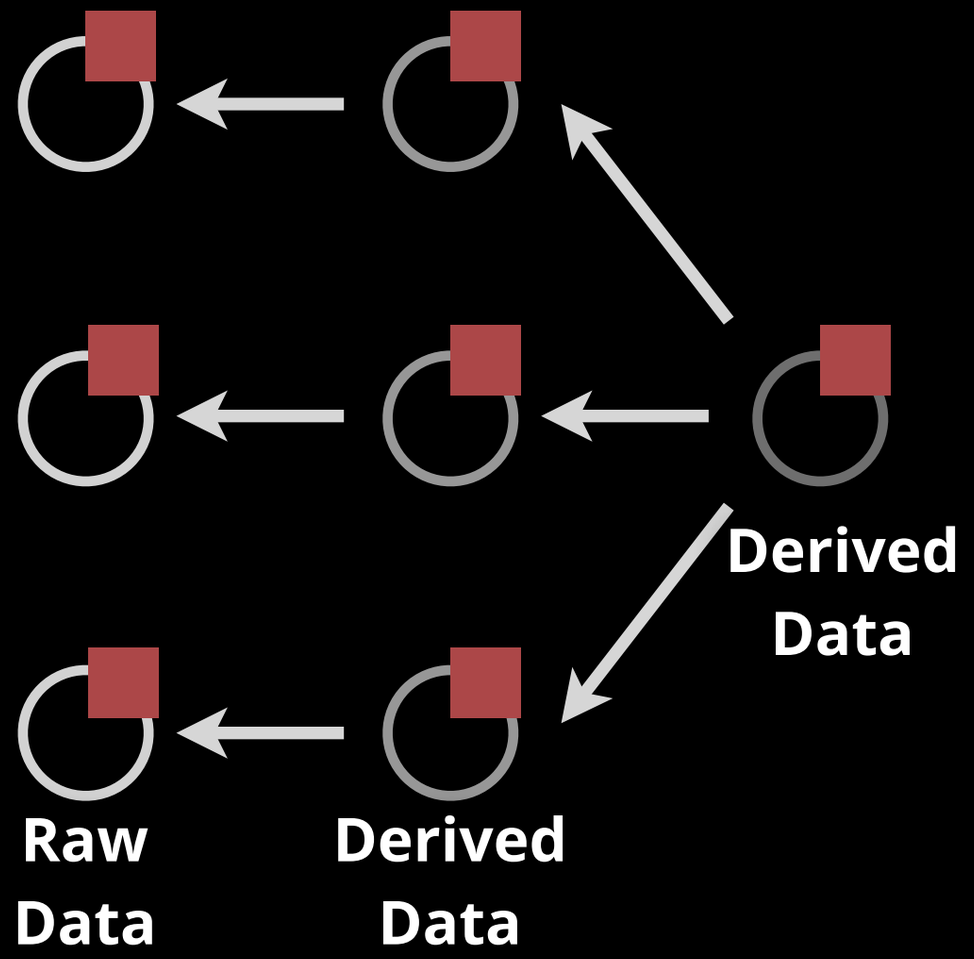
Repository

Meta data annotation



Repository

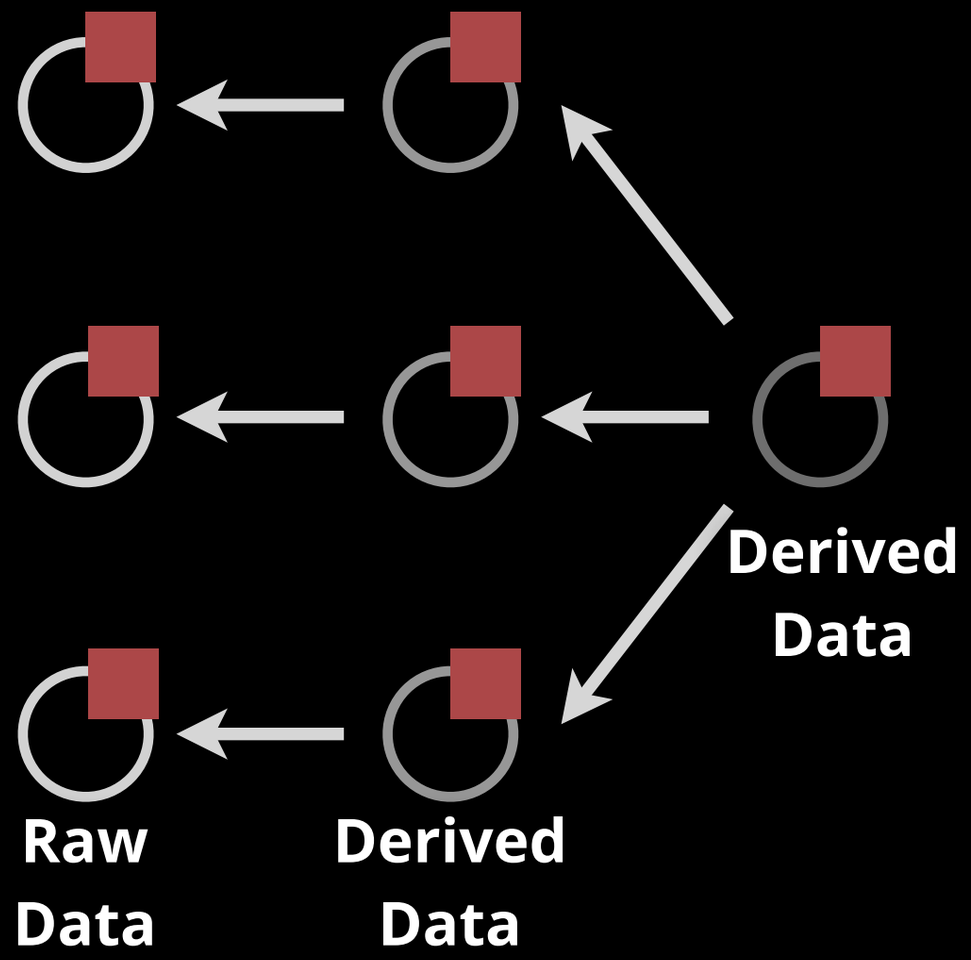
Meta data annotation



Repository

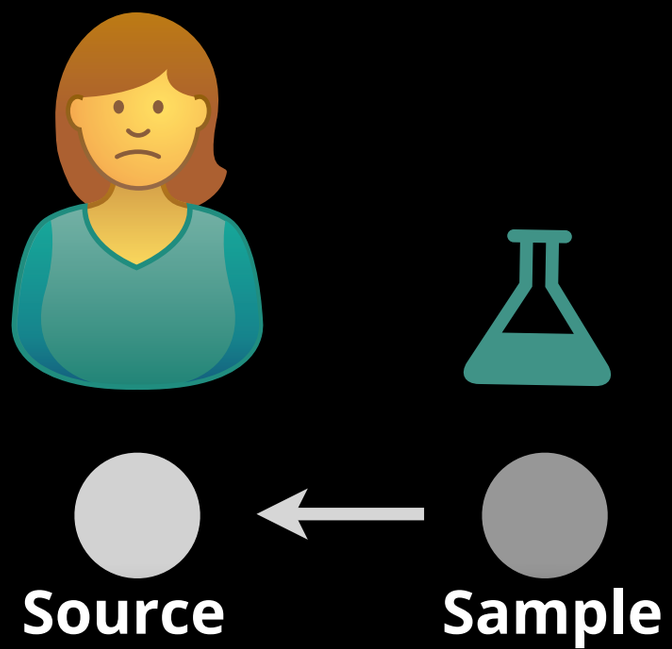
Meta data annotation

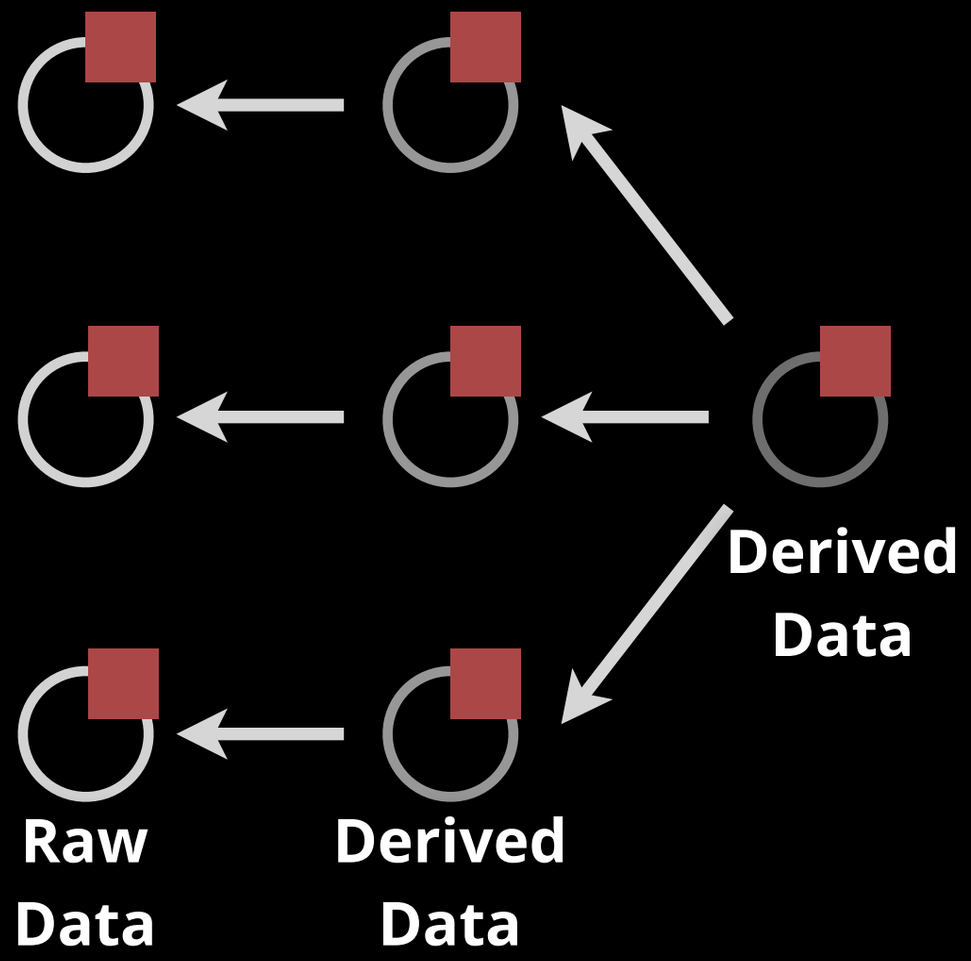
Source



Repository

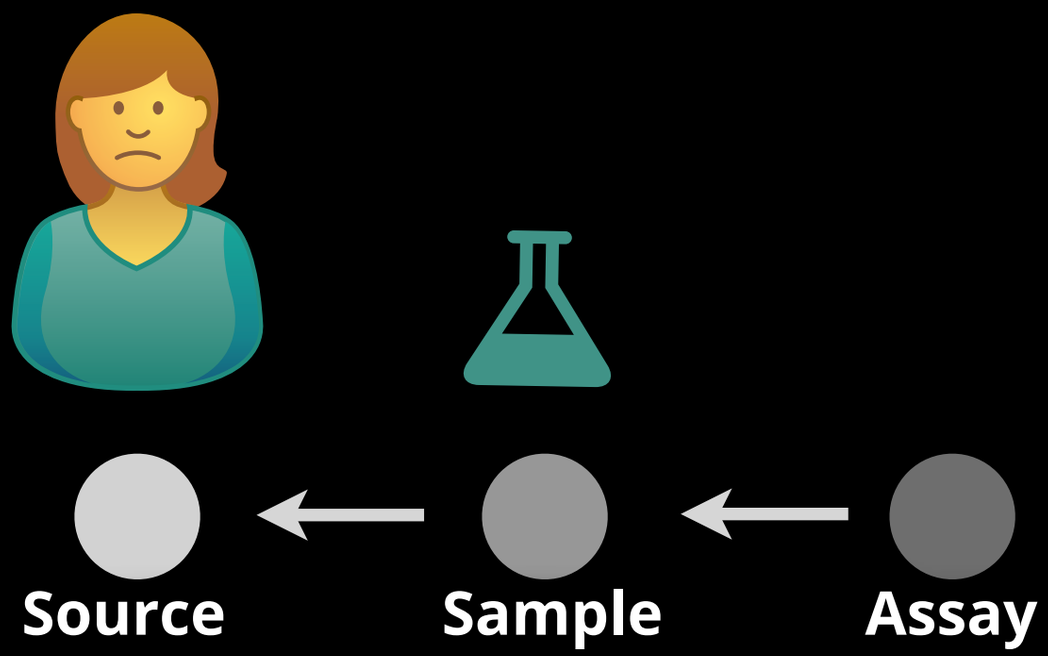
Meta data annotation

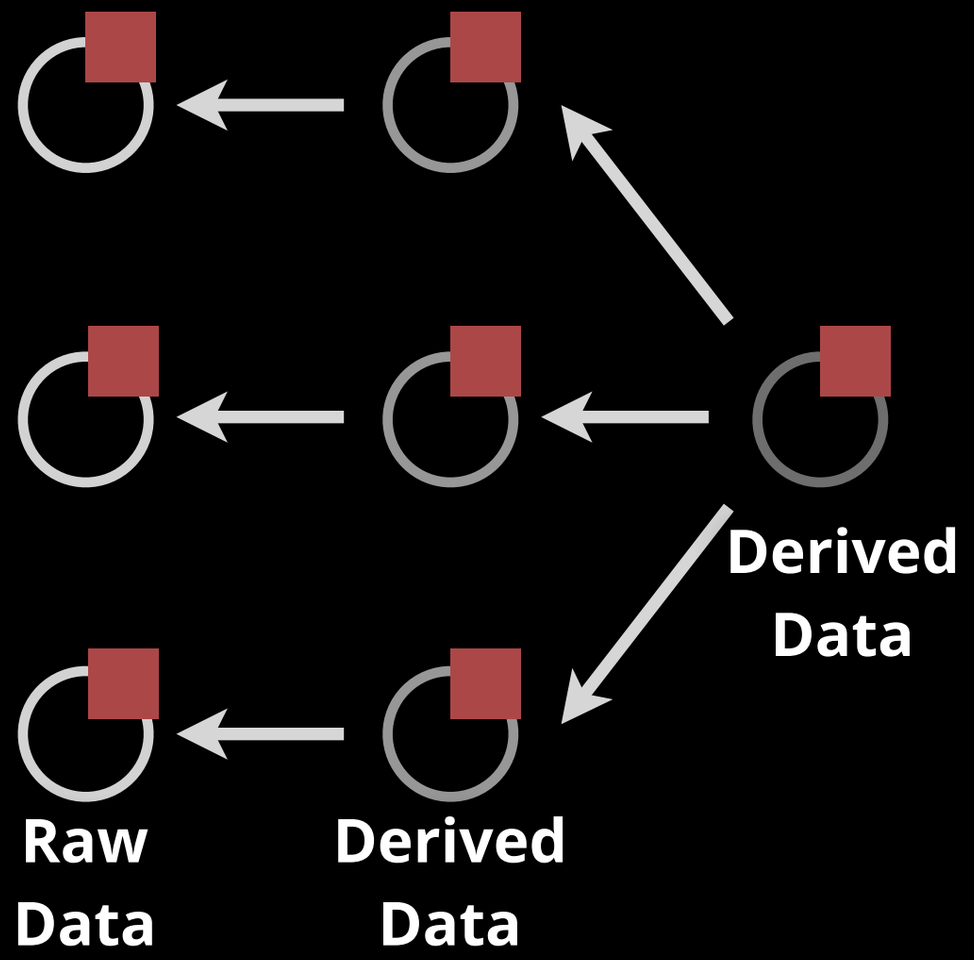




Repository

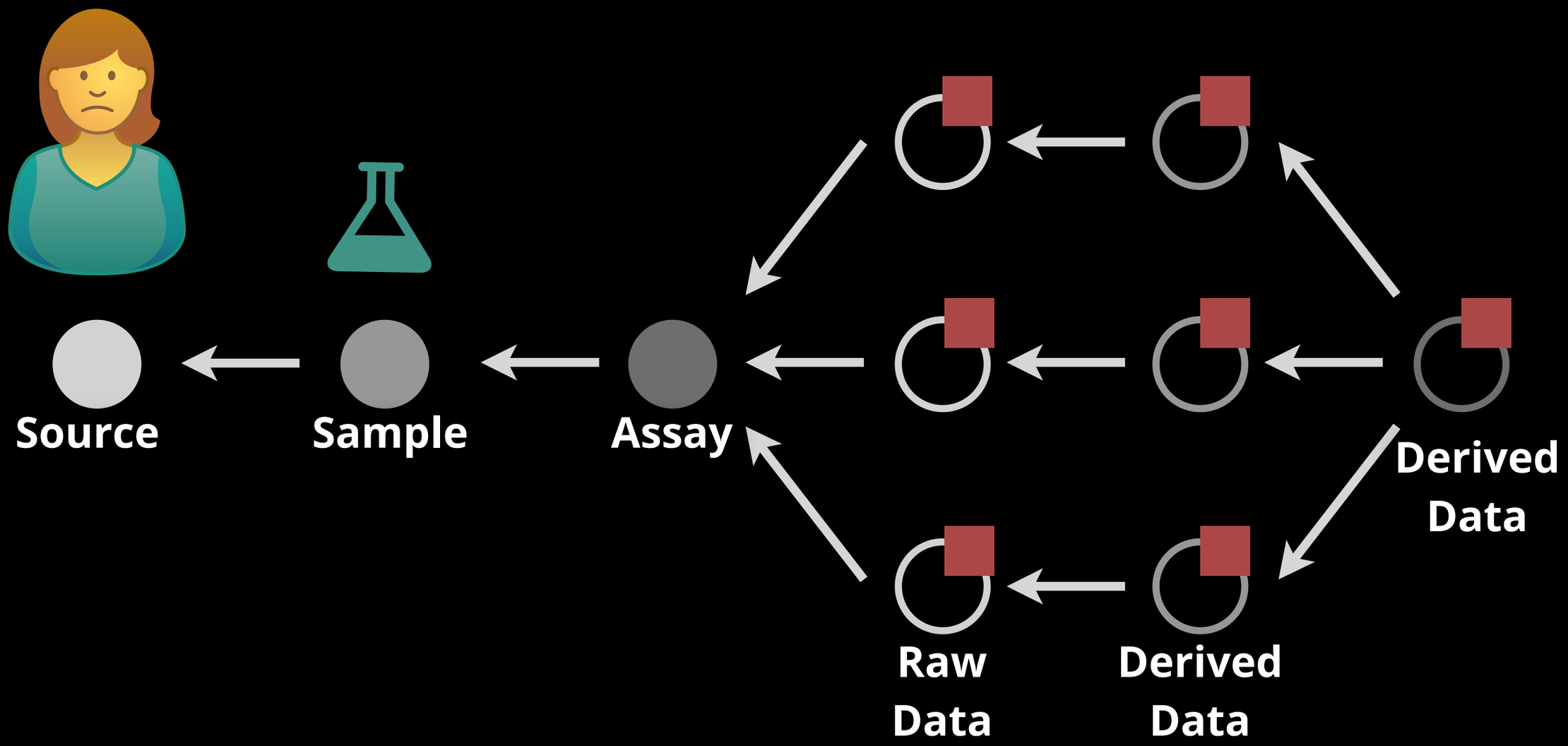
Meta data annotation





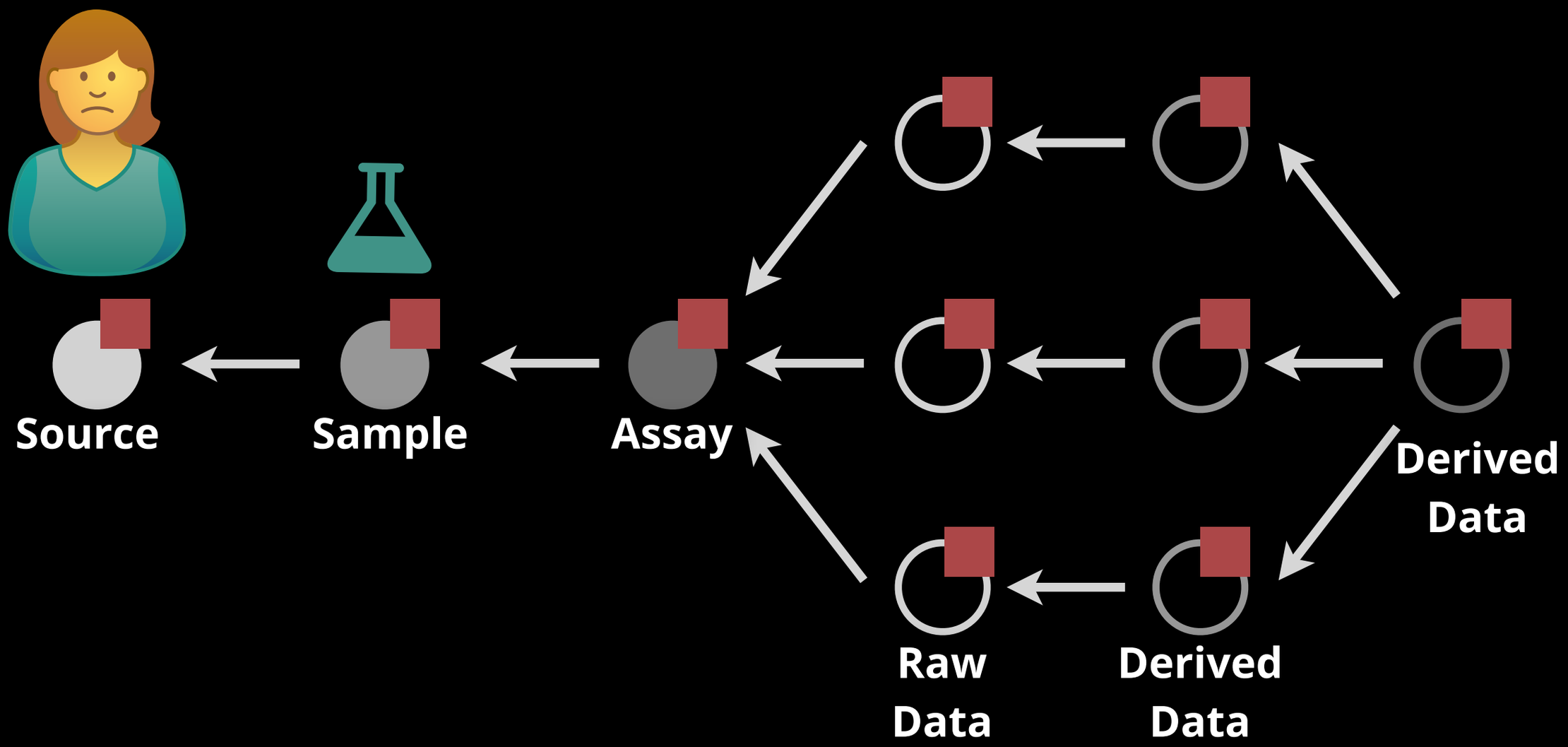
Repository

Meta data annotation



Repository

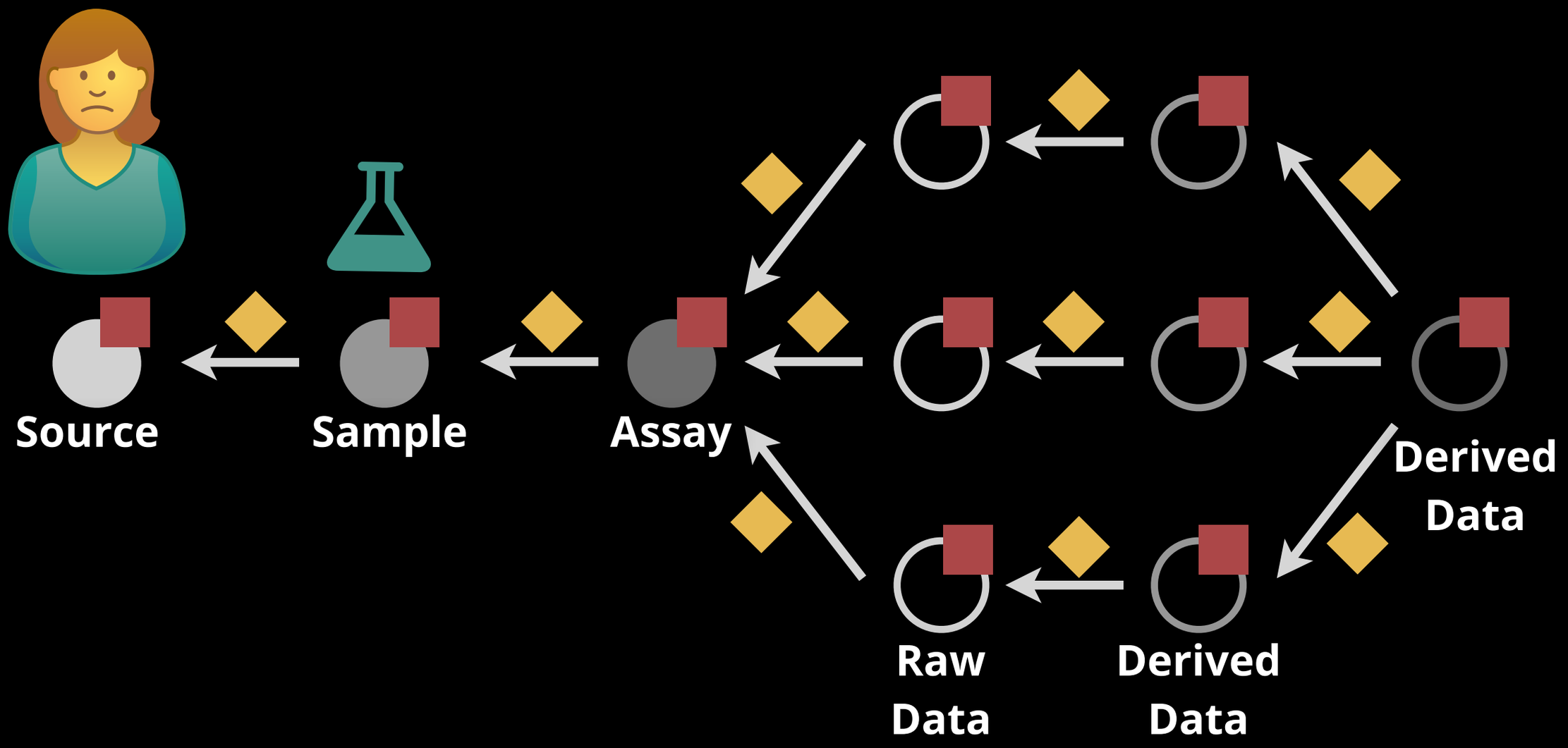
Meta data annotation



Repository

Meta data annotation

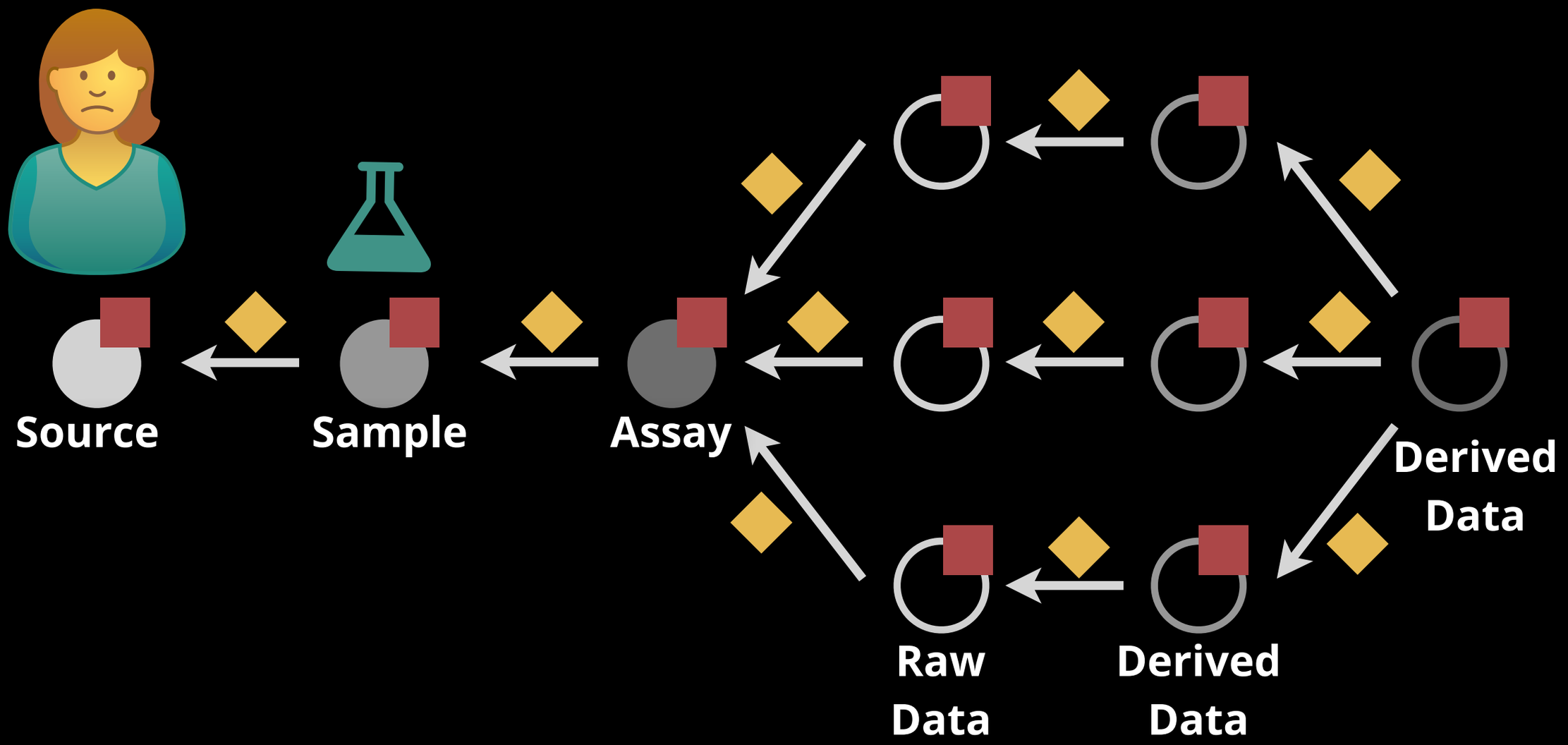
Provenance tracking



Repository

Meta data annotation
Provenance tracking

ISA-Tab



Repository

Meta data annotation
Provenance tracking

ISA-Tab

Meta data and provenance import: ISA-Tab

Repository

Meta data annotation
Provenance tracking

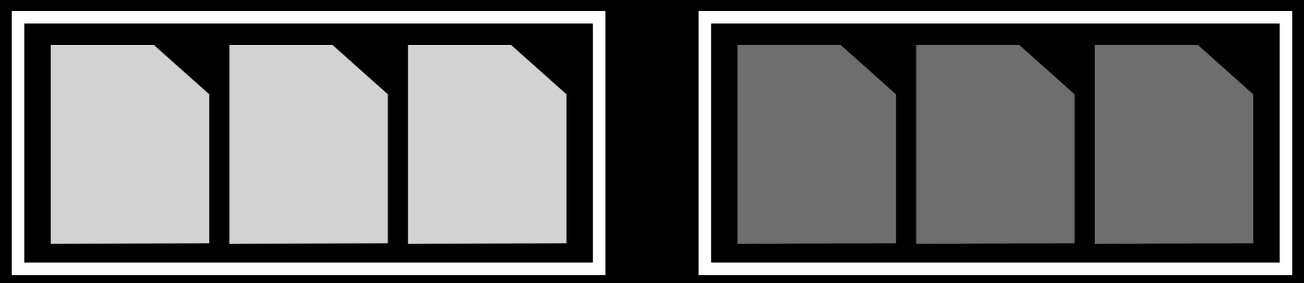
ISA-Tab

Meta data and provenance import: ISA-Tab
Meta data import: tabular files

Data stored locally
Data stored remotely: download on demand

Repository

Sample Browsing



Refinery – Data Set ENCODE-X Chromatin

127.0.0.1:8000/data_sets/672a9d11-666a-11e2-965f-c42c031a67c4/

Refinery | Home | Statistics | About | Nils Gehlenborg | Logout

Data Set ENCODE-X Chromatin

Samples | Attributes | Analyses | Downloads | Details | Sharing

1328 of 1328 | View in IGV | Showing: Data, Annotation | Selection: Save, Load | Workflow

Facets

- Cell Or Tissue
- Platform
- Species
- Factor Function
- Factor
- Treatment
- Data Source
- Genome Build
- Antibody
- Data Type
- Lab

Reset All

List | Matrix

Columns | 10 | 20 | 50 | 100 | 500

Cell Or Tissue	Platform	Species	Factor Function	Factor	Data Source	Antibody	Data Type	Lab
A549	Illumina sequencing	H. sapiens	Histone modification active	H3K4me3	ENCODE	H3K4me3	Histone ChIP-seq	Broad
A549	Illumina sequencing	H. sapiens	Histone modification active	H3K79me2	ENCODE	H3K79me2	Histone ChIP-seq	Broad
A549	Illumina sequencing	H. sapiens	Histone modification active	H3K9ac	ENCODE	H3K9ac	Histone ChIP-seq	Broad
A549	Illumina sequencing	H. sapiens	Histone modification active	H3K4me3	ENCODE	H3K4me3	Histone ChIP-seq	UW
A549	Illumina sequencing	H. sapiens	Histone modification active	H3K4me3	ENCODE	H3K4me3	Histone ChIP-seq	Broad
A549	Illumina sequencing	H. sapiens	Enhancer associated	CEBPB	ENCODE	CEBPB	non-Histone ChIP-seq	Stanford
A549	Illumina sequencing	H. sapiens	Transcription	PolII	ENCODE	Pol2	non-Histone ChIP-seq	HudsonAlpha
A549	Illumina sequencing	H. sapiens	Transcription	PolII	ENCODE	Pol2	non-Histone ChIP-seq	HudsonAlpha
A549	Illumina sequencing	H. sapiens	Insulator	CTCF	ENCODE	CTCF	non-Histone ChIP-seq	UW
A549	Illumina sequencing	H. sapiens	Enhancer associated	P300	ENCODE	p300	non-Histone ChIP-seq	HudsonAlpha
A549	Illumina sequencing	H. sapiens	Transcription	PolII	ENCODE	Pol2(phosphoS2)	non-Histone ChIP-seq	Stanford
A549	Illumina sequencing	H. sapiens	Transcription	PolII	ENCODE	Pol2	non-Histone ChIP-seq	UT-A
A549	Illumina sequencing	H. sapiens	Insulator	CTCF	ENCODE	CTCF	non-Histone ChIP-seq	UT-A
A549	Illumina sequencing	H. sapiens	Insulator	CTCF	ENCODE	CTCF_(SC-5916)	non-Histone ChIP-seq	HudsonAlpha
A549	Illumina sequencing	H. sapiens	Insulator	CTCF	ENCODE	CTCF_(SC-5916)	non-Histone ChIP-seq	HudsonAlpha
A549	Illumina sequencing	H. sapiens	Histone modification active	H3K4me3	ENCODE	H3K4me3	Histone ChIP-seq	Broad
A549	Illumina sequencing	H. sapiens	Histone modification active	H3K79me2	ENCODE	H3K79me2	Histone ChIP-seq	Broad
A549	Illumina sequencing	H. sapiens	Histone modification active	H3K4me3	ENCODE	H3K4me3	Histone ChIP-seq	Broad

Refinery – Data Set ENCODE-X Chromatin

127.0.0.1:8000/data_sets/672a9d11-666a-11e2-965f-c42c031a67c4/

Refinery | Home | Statistics | About | Nils Gehlenborg | Logout

Data Set ENCODE-X Chromatin

Samples | Attributes | Analyses | Downloads | Details | Sharing

1328 of 1328 | View in IGV | Showing: Data, Annotation | Selection: Save, Load | Workflow

Facets | List | Matrix

Reset All | Columns | 10 20 50 100 500

Cell Or Tissue

- S2 cell line 124
- Third instart lavae (L3) 124
- BG3 cell line 92
- K562 90
- Lavae stage 3 (L3) 67
- Late embryo 14-16hr (EL) 65
- Kc cell line 63
- Adult head (AH) 54
- Early embryo (EE) 42
- H1-hESC 42
- Early embryo 2-4hr (EE) 38
- HepG2 34
- GM12878 32
- HeLa-S3 28
- Mixed embryo (MXEMB) 27
- HUVEC 24
- K562 + A549 24
- Clone 8 cell line 21
- NHEK 20
- A549 19
- HMEC 16

Cell Or Tissue	Platform	Species	Factor Function	Factor	Data Source	Antibody	Data Type	Lab
A549	Illumina sequencing	H. sapiens	Histone modification active	H3K4me3	ENCODE	H3K4me3	Histone ChIP-seq	Broad
A549	Illumina sequencing	H. sapiens	Histone modification active	H3K79me2	ENCODE	H3K79me2	Histone ChIP-seq	Broad
A549	Illumina sequencing	H. sapiens	Histone modification active	H3K9ac	ENCODE	H3K9ac	Histone ChIP-seq	Broad
A549	Illumina sequencing	H. sapiens	Histone modification active	H3K4me3	ENCODE	H3K4me3	Histone ChIP-seq	UW
A549	Illumina sequencing	H. sapiens	Histone modification active	H3K4me3	ENCODE	H3K4me3	Histone ChIP-seq	Broad
A549	Illumina sequencing	H. sapiens	Enhancer associated	CEBPB	ENCODE	CEBPB	non-Histone ChIP-seq	Stanford
A549	Illumina sequencing	H. sapiens	Transcription	PolII	ENCODE	Pol2	non-Histone ChIP-seq	HudsonAlpha
A549	Illumina sequencing	H. sapiens	Transcription	PolII	ENCODE	Pol2	non-Histone ChIP-seq	HudsonAlpha
A549	Illumina sequencing	H. sapiens	Insulator	CTCF	ENCODE	CTCF	non-Histone ChIP-seq	UW
A549	Illumina sequencing	H. sapiens	Enhancer associated	P300	ENCODE	p300	non-Histone ChIP-seq	HudsonAlpha
A549	Illumina sequencing	H. sapiens	Transcription	PolII	ENCODE	Pol2(phosphoS2)	non-Histone ChIP-seq	Stanford
A549	Illumina sequencing	H. sapiens	Transcription	PolII	ENCODE	Pol2	non-Histone ChIP-seq	UT-A
A549	Illumina sequencing	H. sapiens	Insulator	CTCF	ENCODE	CTCF	non-Histone ChIP-seq	UT-A
A549	Illumina sequencing	H. sapiens	Insulator	CTCF	ENCODE	CTCF_(SC-5916)	non-Histone ChIP-seq	HudsonAlpha
A549	Illumina sequencing	H. sapiens	Insulator	CTCF	ENCODE	CTCF_(SC-5916)	non-Histone ChIP-seq	HudsonAlpha
A549	Illumina sequencing	H. sapiens	Histone modification active	H3K4me3	ENCODE	H3K4me3	Histone ChIP-seq	Broad
A549	Illumina sequencing	H. sapiens	Histone modification active	H3K79me2	ENCODE	H3K79me2	Histone ChIP-seq	Broad
A549	Illumina sequencing	H. sapiens	Histone modification active	H3K4me3	ENCODE	H3K4me3	Histone ChIP-seq	Broad

Refinery – Data Set ENCODE-X Chromatin

127.0.0.1:8000/data_sets/672a9d11-666a-11e2-965f-c42c031a67c4/

Refinery | Home | Statistics | About | Nils Gehlenborg | Logout

Data Set ENCODE-X Chromatin

Samples | Attributes | Analyses | Downloads | Details | Sharing

28 of 1328 | View in IGV | Showing: Data, Annotation | Selection: Save, Load | Workflow

Facets | List | Matrix

Reset All | Columns | 10 20 50 100 500

- Cell Or Tissue
  - K562 28
- Platform
- Species
  - H. sapiens 28
- Factor Function
  - Transcription 13
  - Histone methyltransferase 9
  - Enhancer associated 6
- Factor
- Treatment
- Data Source
- Genome Build
- Antibody
- Data Type
- Lab

Cell Or Tissue	Platform	Species	Factor Function	Factor	Data Source	Antibody	Data Type	Lab
K562	Illumina sequencing	H. sapiens	Transcription	PolIIS5P	Ram et al	RNAPIIS5P	non-Histone ChIP-seq	Bernstein
K562	Illumina sequencing	H. sapiens	Histone methyltransferase	RBBP5	Ram et al	RBBP5	non-Histone ChIP-seq	Bernstein
K562	Illumina sequencing	H. sapiens	Enhancer associated	CBP/CBP1	Ram et al	KAT3A-CBP	non-Histone ChIP-seq	Bernstein
K562	Illumina sequencing	H. sapiens	Enhancer associated	P300	Ram et al	P300	non-Histone ChIP-seq	Bernstein
K562	Illumina sequencing	H. sapiens	Histone methyltransferase	ESET/SETDB1	Ram et al	ESET	non-Histone ChIP-seq	Bernstein
K562	Illumina sequencing	H. sapiens	Histone methyltransferase	RBBP5	ENCODE	RBBP5	non-Histone ChIP-seq	Broad
K562	Illumina sequencing	H. sapiens	Transcription	PolII	ENCODE	Pol2	non-Histone ChIP-seq	Broad
K562	Illumina sequencing	H. sapiens	Enhancer associated	P300	ENCODE	p300	non-Histone ChIP-seq	Broad
K562	Illumina sequencing	H. sapiens	Enhancer associated	CEBPB	ENCODE	CEBPB	non-Histone ChIP-seq	Stanford
K562	Illumina sequencing	H. sapiens	Enhancer associated	CEBPB	ENCODE	CEBPB_(SC-150)	non-Histone ChIP-seq	HudsonAlpha
K562	Illumina sequencing	H. sapiens	Histone methyltransferase	ESET/SETDB1	ENCODE	SETDB1	non-Histone ChIP-seq	USC
K562	Illumina sequencing	H. sapiens	Histone methyltransferase	ESET/SETDB1	ENCODE	SETDB1	non-Histone ChIP-seq	USC
K562	Illumina sequencing	H. sapiens	Enhancer associated	P300	ENCODE	p300	non-Histone ChIP-seq	Stanford
K562	Illumina sequencing	H. sapiens	Transcription	PolII	ENCODE	Pol2(phosphoS2)	non-Histone ChIP-seq	Stanford
K562	Illumina sequencing	H. sapiens	Transcription	PolII	ENCODE	Pol2	non-Histone ChIP-seq	Stanford
K562	Illumina sequencing	H. sapiens	Transcription	PolII	ENCODE	Pol2(phosphoS2)	non-Histone ChIP-seq	Stanford
K562	Illumina sequencing	H. sapiens	Transcription	PolII	ENCODE	Pol2	non-Histone ChIP-seq	Yale
K562	Illumina sequencing	H. sapiens	Transcription	PolII	ENCODE	Pol2	non-Histone ChIP-seq	Stanford

Refinery – Data Set ENCODE-X Chromatin

127.0.0.1:8000/data_sets/672a9d11-666a-11e2-965f-c42c031a67c4/

Refinery | Home | Statistics | About | Nils Gehlenborg | Logout

Data Set ENCODE-X Chromatin

Samples | Attributes | Analyses | Downloads | Details | Sharing

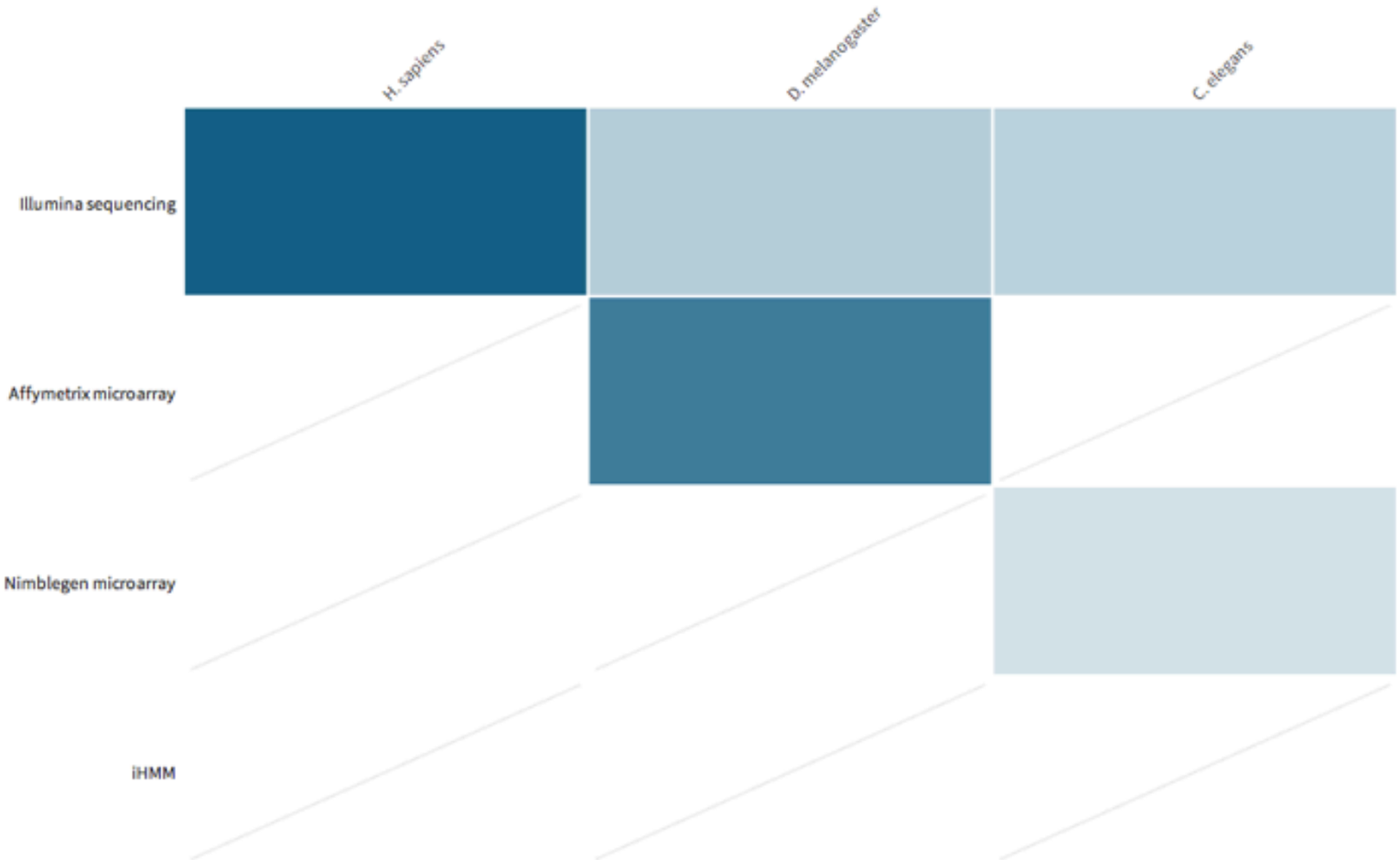
1328 of 1328 | View in IGV | Showing: Data, Annotation | Selection: Save, Load | Workflow

Facets | List | Matrix

Reset All

Rows: Platform (4) | Columns: Species (3)

- Cell Or Tissue
- Platform
- Species
- Factor Function
- Factor
- Treatment
- Data Source
- Genome Build
- Antibody
- Data Type
- Lab



Refinery – Data Set ENCODE-X Chromatin

127.0.0.1:8000/data_sets/672a9d11-666a-11e2-965f-c42c031a67c4/

Refinery | Home | Statistics | About | Nils Gehlenborg | Logout

Data Set ENCODE-X Chromatin

Samples | Attributes | Analyses | Downloads | Details | Sharing

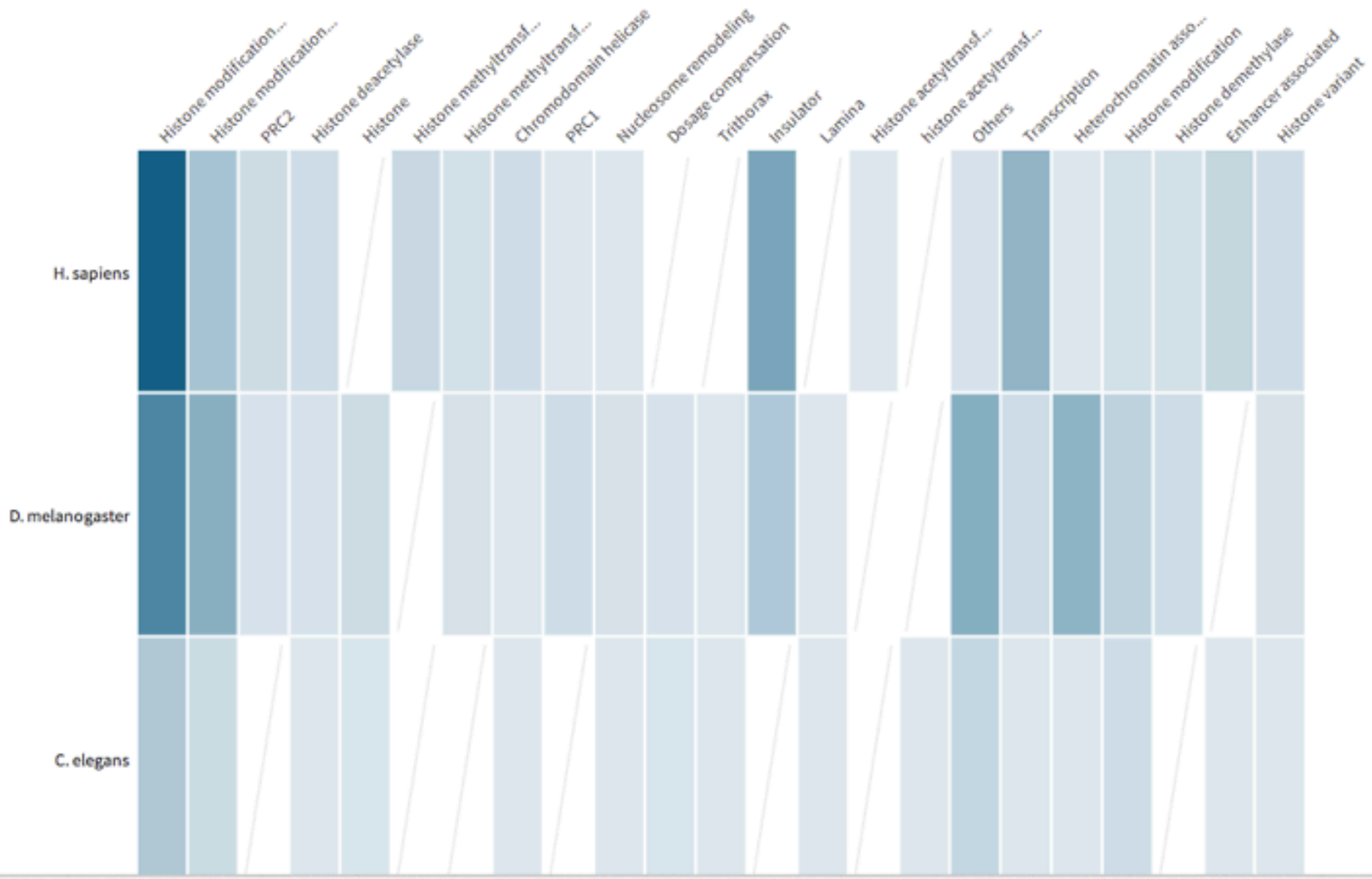
1328 of 1328 | View in IGV | Showing: Data, Annotation | Selection: Save, Load | Workflow

Facets | List | Matrix

Reset All

Rows: Species (3) | Columns: Factor Function (23)

- Cell Or Tissue
- Platform
- Species
- Factor Function
- Factor
- Treatment
- Data Source
- Genome Build
- Antibody
- Data Type
- Lab



Refinery – Data Set ENCODE-X Chromatin

127.0.0.1:8000/data_sets/672a9d11-666a-11e2-965f-c42c031a67c4/

Reader

Refinery Home Statistics About Nils Gehlenborg | Logout

Data Set ENCODE-X Chromatin

Samples | Attributes | Analyses | Downloads | Details | Sharing

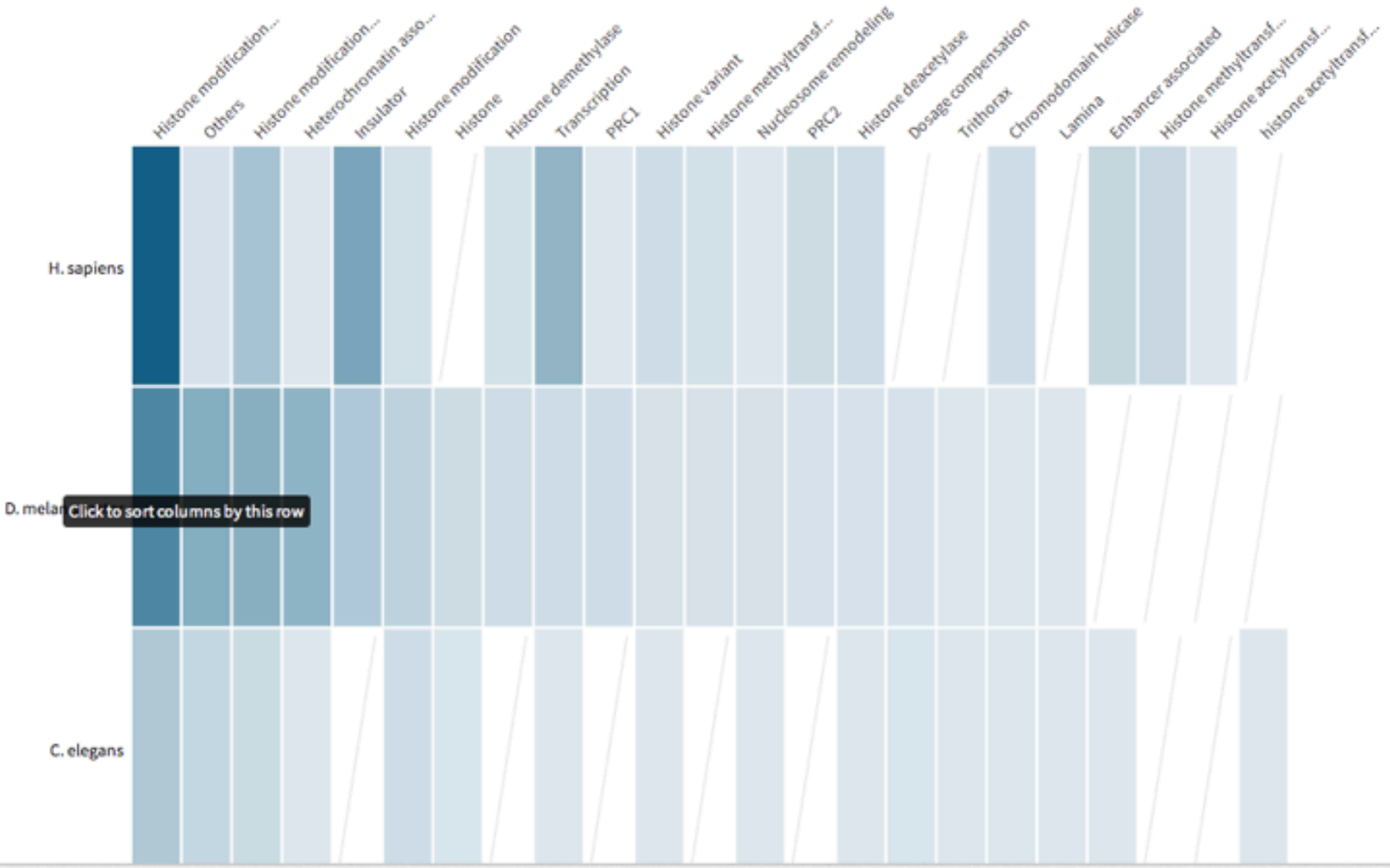
1328 of 1328 View in IGV Showing Data Annotation Selection Save Load Workflow

Facets List Matrix

Reset All

Rows Species (3) Columns Factor Function (23)

- Cell Or Tissue
- Platform
- Species
- Factor Function
- Factor
- Treatment
- Data Source
- Genome Build
- Antibody
- Data Type
- Lab



Refinery – Data Set ENCODE-X Chromatin

127.0.0.1:8000/data_sets/672a9d11-666a-11e2-965f-c42c031a67c4/

Refinery | Home | Statistics | About | Nils Gehlenborg | Logout

Data Set ENCODE-X Chromatin

Samples | Attributes | Analyses | Downloads | Details | Sharing

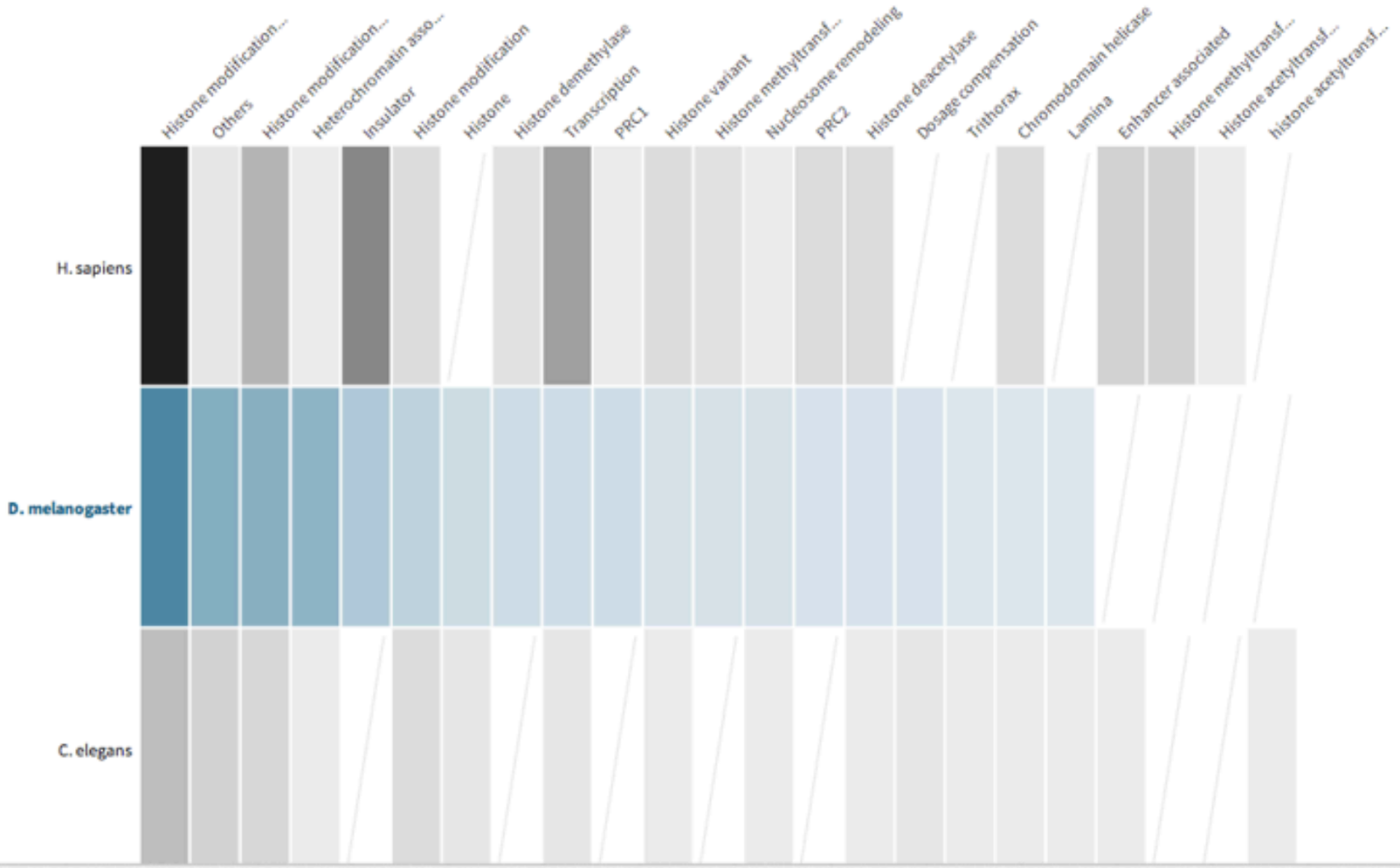
593 of 1328 | View in IGV | Showing: Data, Annotation | Selection: Save, Load | Workflow

Facets | List | Matrix

Reset All

Rows: Species (3) | Columns: Factor Function (23)

- Cell Or Tissue
- Platform
- Species
  - D. melanogaster 593
- Factor Function
- Factor
- Treatment
- Data Source
- Genome Build
- Antibody
- Data Type
- Lab



Refinery – Data Set ENCODE-X Chromatin

127.0.0.1:8000/data_sets/672a9d11-666a-11e2-965f-c42c031a67c4/

Reader

Refinery Home Statistics About — Nils Gehlenborg | Logout

Data Set ENCODE-X Chromatin

Samples | Attributes | Analyses | Downloads | Details | Sharing

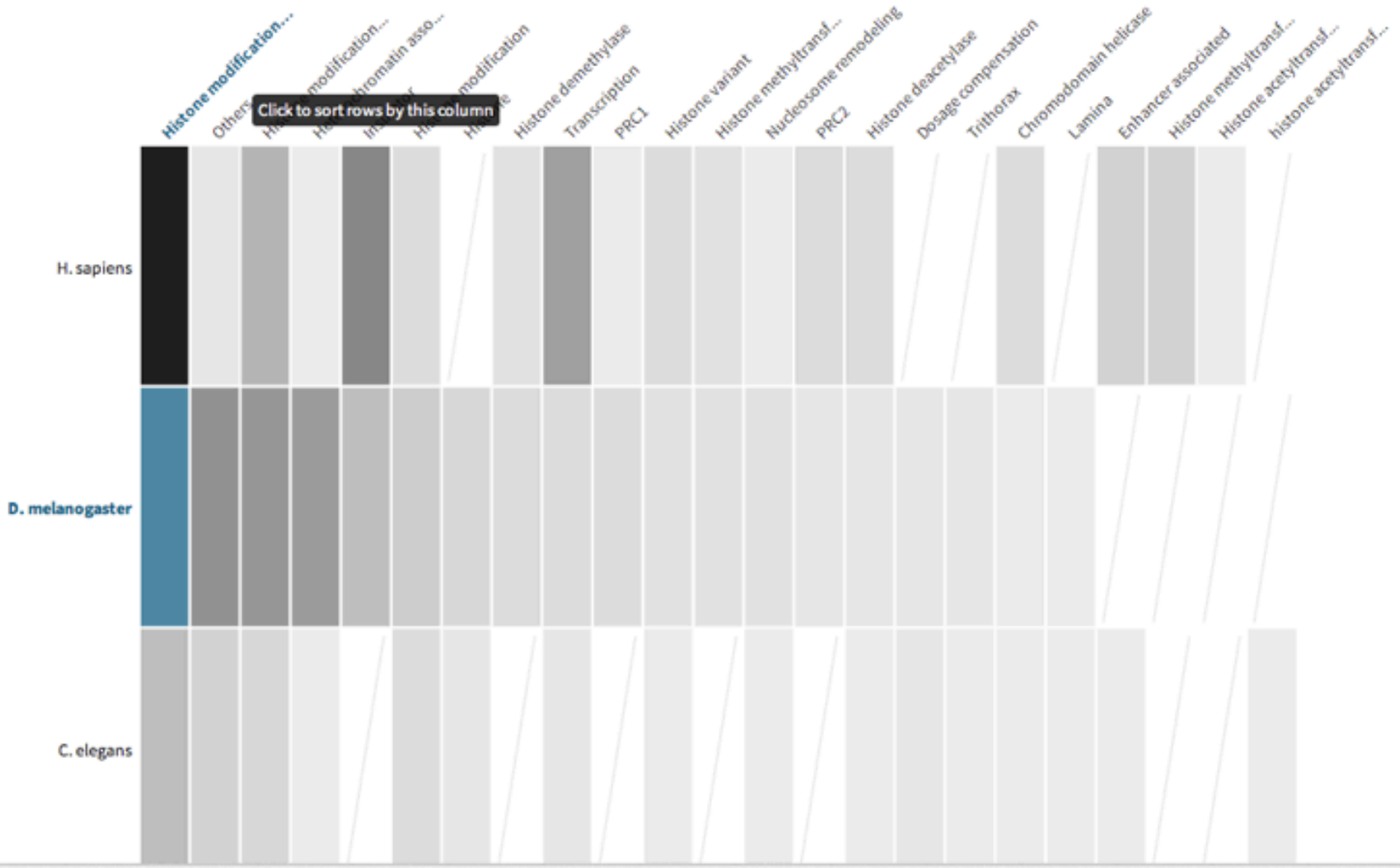
143 of 1328 | View in IGV | Showing: Data, Annotation | Selection: Save, Load | Workflow

Facets | List | Matrix

Reset All

Rows: Species (3) | Columns: Factor Function (23)

- Cell Or Tissue
- Platform
- Species
  - D. melanogaster 143
- Factor Function
  - Histone modification active 143
- Factor
- Treatment
- Data Source
- Genome Build
- Antibody
- Data Type
- Lab



Refinery – Data Set ENCODE-X Chromatin

127.0.0.1:8000/data_sets/672a9d11-666a-11e2-965f-c42c031a67c4/

Refinery | Home | Statistics | About | Nils Gehlenborg | Logout

Data Set ENCODE-X Chromatin

Samples | Attributes | Analyses | Downloads | Details | Sharing

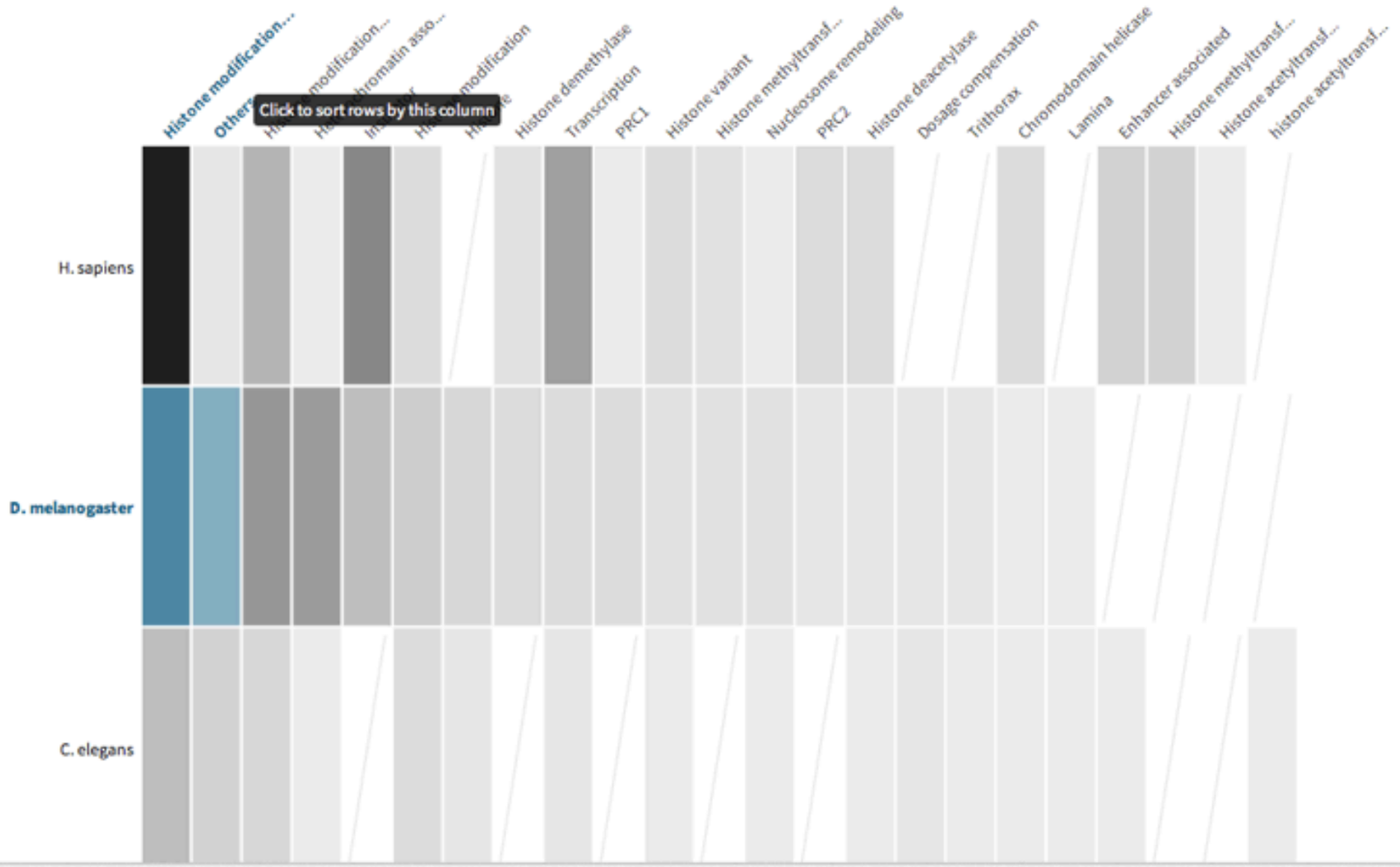
231 of 1328 | View in IGV | Showing: Data, Annotation | Selection: Save, Load | Workflow

Facets | List | Matrix

Reset All

Rows: Species (3) | Columns: Factor Function (23)

- Cell Or Tissue
- Platform
- Species
  - D. melanogaster 231
- Factor Function
  - Histone modification active 143
  - Others 88
- Factor
- Treatment
- Data Source
- Genome Build
- Antibody
- Data Type
- Lab



Refinery – Data Set ENCODE-X Chromatin

127.0.0.1:8000/data_sets/672a9d11-666a-11e2-965f-c42c031a67c4/

Refinery | Home | Statistics | About | Nils Gehlenborg | Logout

Data Set ENCODE-X Chromatin

Samples | Attributes | Analyses | Downloads | Details | Sharing

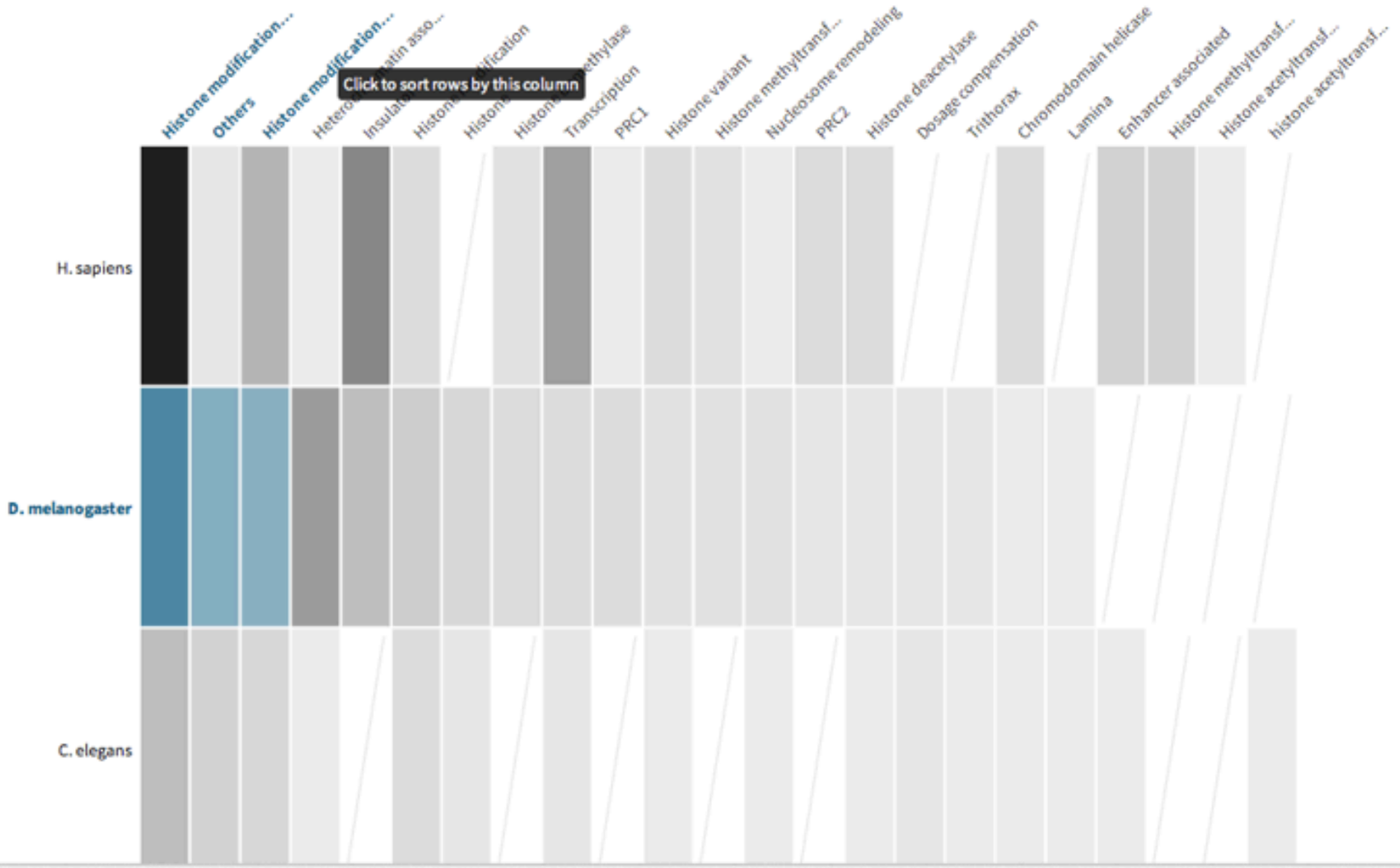
314 of 1328 | View in IGV | Showing: Data, Annotation | Selection: Save, Load | Workflow

Facets | List | Matrix

Reset All

Rows: Species (3) | Columns: Factor Function (23)

- Cell Or Tissue
- Platform
- Species
  - D. melanogaster 314
- Factor Function
  - Histone modification active 143
  - Others 88
  - Histone modification repressive 83
- Factor
- Treatment
- Data Source
- Genome Build
- Antibody
- Data Type
- Lab



Refinery – Data Set ENCODE-X Chromatin

127.0.0.1:8000/data_sets/672a9d11-666a-11e2-965f-c42c031a67c4/

Reader

Refinery Home Statistics About — Nils Gehlenborg | Logout

Data Set ENCODE-X Chromatin

Samples | Attributes | Analyses | Downloads | Details | Sharing

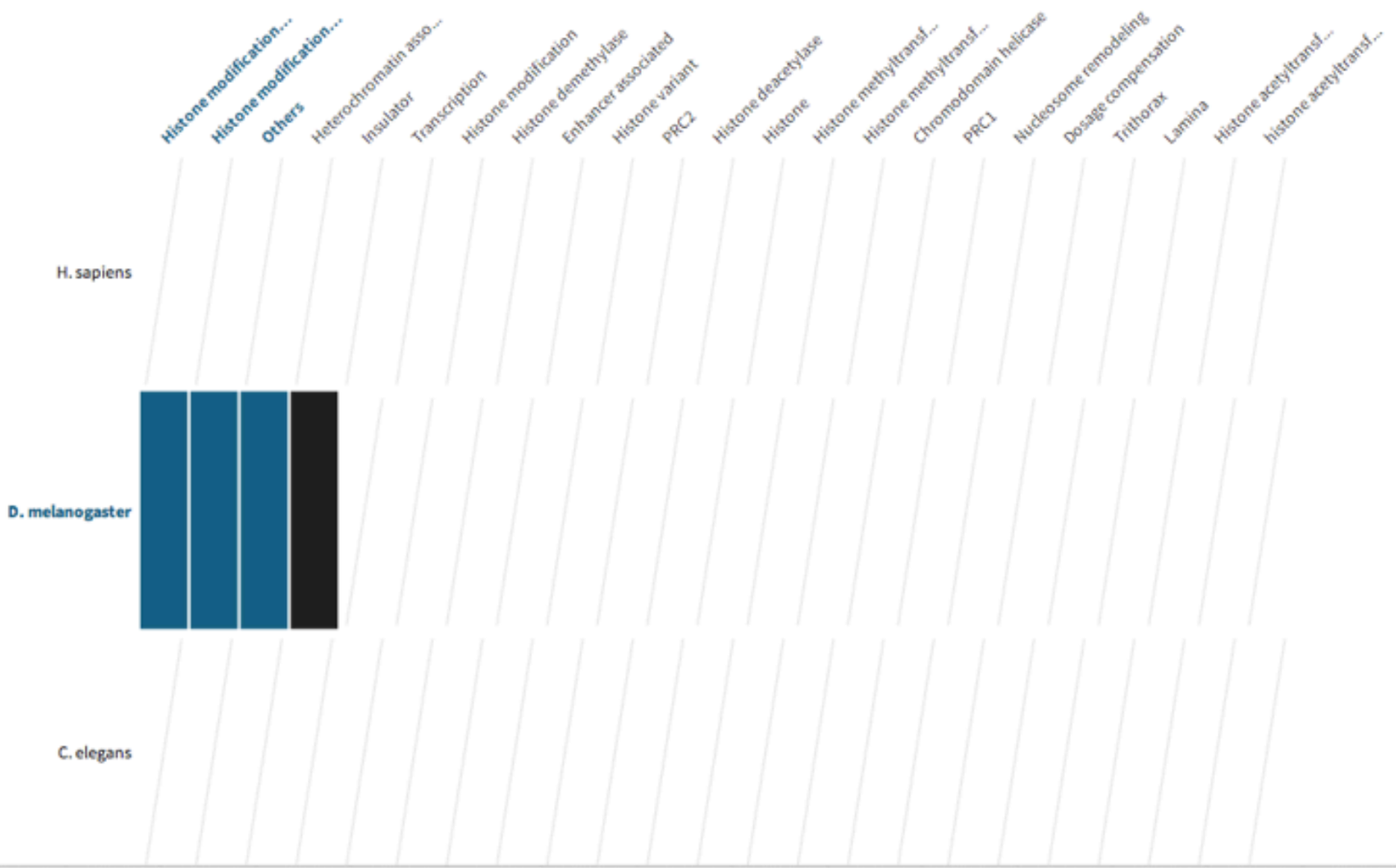
3 of 1328 | View in IGV | Showing: Data, Annotation | Selection: Save, Load | Workflow

Facets | List | Matrix

Reset All

Rows: Species (3) | Columns: Factor Function (23)

- Cell Or Tissue
  - L3 Sexed Male 3
- Platform
- Species
  - D. melanogaster 3
- Factor Function
  - Histone modification active 1
  - Histone modification repressive 1
  - Others 1
- Factor
- Treatment
- Data Source
- Genome Build
- Antibody
- Data Type
- Lab



Workflow Engine

Galaxy

Tools
Toolshed
Workflow Editor
API

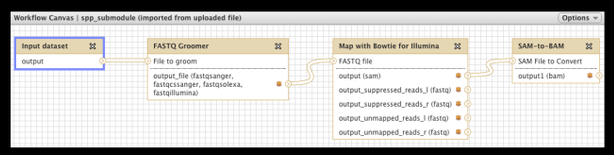
Workflow Engine

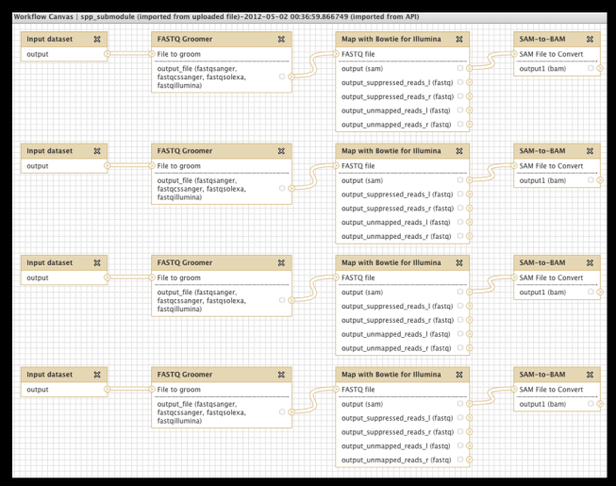
Galaxy

Workflow Template

API

Workflow Instance





Workflow Engine

Galaxy

Workflow Inputs

API

Workflow Outputs

Refinery – Data Set Test 1: Request for Comments (RFC) Test

127.0.0.1:8000/data_sets/5321458a-c49a-11e2-83ae-c42c031a67c4/

Refinery | Home | Statistics | About | Nils Gehlenborg | Logout

Data Set Test 1: Request for Comments (RFC) Test

Samples | Attributes | Analyses | Downloads | Details | Sharing

23 of 23 | View in IGV | Showing: Data, Annotation | Selection: Save, Load | Workflow

Facets | List | Matrix

Reset All | Columns | 10 20 50 100 500

- Author
- Month
- Year
- Type
- File Type
- Analysis

Title	Author	Month	Year	Type	File Type	Name
Device independent graphical display description	McConnell	June	1971	Raw Data File	txt	http://gehlenborg.com/wp-content/uploads/rfc/rfc177.txt
Graphics Facilities at Ames Research Center	McConnell	April	1971	Raw Data File	txt	http://gehlenborg.com/wp-content/uploads/rfc/rfc126.txt
Network Graphics meeting	Vezza	April	1971	Raw Data File	txt	http://gehlenborg.com/wp-content/uploads/rfc/rfc134.txt
Network graphic attention handling	Cotton	June	1971	Raw Data File	txt	http://gehlenborg.com/wp-content/uploads/rfc/rfc178.txt
Pressure from the Chairman	Crocker	March	1971	Derived Data File	txt	1_test tool out
Pressure from the Chairman	Crocker	March	1971	Derived Data File	txt	1_test_02
Pressure from the Chairman	Crocker	March	1971	Derived Data File	txt	1_test_04
Pressure from the Chairman	Crocker	March	1971	Raw Data File	txt	http://gehlenborg.com/wp-content/uploads/rfc/rfc111.txt
Pressure from the Chairman	Crocker	March	1971	Derived Data File	txt	1_test tool out
Pressure from the Chairman	Crocker	March	1971	Derived Data File	txt	1_test tool out
Pressure from the Chairman	Crocker	March	1971	Derived Data File	txt	1_test_03
Pressure from the Chairman	Crocker	March	1971	Derived Data File	txt	1_test_01
Proposal for a Network Standard Format for a Data Stream to Control	Crocker	January	1971	Raw Data File	txt	http://gehlenborg.com/wp-

Refinery – Data Set Test 1: Request for Comments (RFC) Test

127.0.0.1:8000/data_sets/5321458a-c49a-11e2-83ae-c42c031a67c4/

Refinery | Home | Statistics | About | Nils Gehlenborg | Logout

Data Set Test 1: Request for Comments (RFC) Test

Samples | Attributes | Analyses | Downloads | Details | Sharing

23 of 23 | View in IGV | Showing: Data, Annotation | Selection: Save, Load | Workflow

Test workflow: SPP analog

Test workflow: 5 steps without branching

Facets | List | Matrix

Reset All | Columns | 10 20 50 100 500

- Author
- Month
- Year
- Type
- File Type
- Analysis

Title	Author	Month	Year	Type	File Type	Name
Device independent graphical display description	McConnell	June	1971	Raw Data File	txt	http://gehlenborg.com/wp-content/uploads/rfc/rfc177.txt
Graphics Facilities at Ames Research Center	McConnell	April	1971	Raw Data File	txt	http://gehlenborg.com/wp-content/uploads/rfc/rfc126.txt
Network Graphics meeting	Vezza	April	1971	Raw Data File	txt	http://gehlenborg.com/wp-content/uploads/rfc/rfc134.txt
Network graphic attention handling	Cotton	June	1971	Raw Data File	txt	http://gehlenborg.com/wp-content/uploads/rfc/rfc178.txt
Pressure from the Chairman	Crocker	March	1971	Derived Data File	txt	1_test tool out
Pressure from the Chairman	Crocker	March	1971	Derived Data File	txt	1_test_02
Pressure from the Chairman	Crocker	March	1971	Derived Data File	txt	1_test_04
Pressure from the Chairman	Crocker	March	1971	Raw Data File	txt	http://gehlenborg.com/wp-content/uploads/rfc/rfc111.txt
Pressure from the Chairman	Crocker	March	1971	Derived Data File	txt	1_test tool out
Pressure from the Chairman	Crocker	March	1971	Derived Data File	txt	1_test tool out
Pressure from the Chairman	Crocker	March	1971	Derived Data File	txt	1_test_03
Pressure from the Chairman	Crocker	March	1971	Derived Data File	txt	1_test_01
Proposal for a Network Standard Format for a Data Stream to Control	Crocker	January	1971	Raw Data File	txt	http://gehlenborg.com/wp-

Refinery – Data Set Test 1: Request for Comments (RFC) Test

127.0.0.1:8000/data_sets/5321458a-c49a-11e2-83ae-c42c031a67c4/

Refinery | Home | Statistics | About | Nils Gehlenborg | Logout

Data Set Test 1: Request for Comments (RFC) Test

Samples | Attributes | Analyses | Downloads | Details | Sharing

23 of 23 | View in IGV | Showing: Data, Annotation | Selection: Save, Load | Workflow: Test workflow: 5 steps without branching

Facets | List | Matrix

Reset All | Columns | 10 20 50 100 500

- Author
- Month
- Year
- Type
- File Type
- Analysis

Title	Author	Month	Year	Type	File Type	Name
Device independent graphical display description	McConnell	June	1971	Raw Data File	txt	http://gehlenborg.com/wp-content/uploads/rfc/rfc177.txt
Graphics Facilities at Ames Research Center	McConnell	April	1971	Raw Data File	txt	http://gehlenborg.com/wp-content/uploads/rfc/rfc126.txt
Network Graphics meeting	Vezza	April	1971	Raw Data File	txt	http://gehlenborg.com/wp-content/uploads/rfc/rfc134.txt
Network graphic attention handling	Cotton	June	1971	Raw Data File	txt	http://gehlenborg.com/wp-content/uploads/rfc/rfc178.txt
Pressure from the Chairman	Crocker	March	1971	Derived Data File	txt	1_test tool out
Pressure from the Chairman	Crocker	March	1971	Derived Data File	txt	1_test_02
Pressure from the Chairman	Crocker	March	1971	Derived Data File	txt	1_test_04
Pressure from the Chairman	Crocker	March	1971	Raw Data File	txt	http://gehlenborg.com/wp-content/uploads/rfc/rfc111.txt
Pressure from the Chairman	Crocker	March	1971	Derived Data File	txt	1_test tool out
Pressure from the Chairman	Crocker	March	1971	Derived Data File	txt	1_test tool out
Pressure from the Chairman	Crocker	March	1971	Derived Data File	txt	1_test_03
Pressure from the Chairman	Crocker	March	1971	Derived Data File	txt	1_test_01
Proposal for a Network Standard Format for a Data Stream	Crocker	January	1971	Raw Data	txt	http://gehlenborg.com/wp-

Run Analysis

Set1

Crocker after

Run Analysis

Refinery

127.0.0.1:8000/analysis_manager/bf5ff2ee-f043-11e2-8364-c42c031a67c4/

Reader

Refinery Home Statistics About Nils Gehlenborg | Logout

Analysis Test workflow: 5 steps without branching 2013-07-19 @ 03:20:53

Status

Test workflow: 5 steps without branching started on July 19, 2013, 3:20 a.m.. (View in Galaxy)

Preprocessing

Finished File upload is complete.

Execution

0%

Postprocessing

Waiting File download is pending.

Refinery – Data Set Test 1: Request for Comments (RFC) Test

127.0.0.1:8000/data_sets/5321458a-c49a-11e2-83ae-c42c031a67c4/

Reader

Refinery Home Statistics About Nils Gehlenborg | Logout

Data Set Test 1: Request for Comments (RFC) Test

Samples | Attributes | Analyses | Downloads | Details | Sharing

2 of 23 View in IGV Showing Data Annotation Selection Save Load Workflow

Facets

Reset All

- Author
- Month
- Year
- Type
- File Type

- Analysis
  - Test workflow: 5 steps without branching 2013-06- 2 20 @ 13:32:08

List Matrix

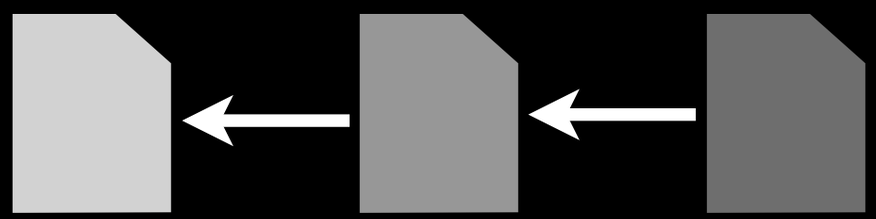
Columns 10 20 50 100 500

Title	Author	Month	Year	Type	File Type	Name
Pressure from the Chairman	Crocker	March	1971	Derived Data File	txt	1_test tool out
Proposal for a Network Standard Format for a Data Stream to Control Graphics Display	Crocker	January	1971	Derived Data File	txt	2_test tool out

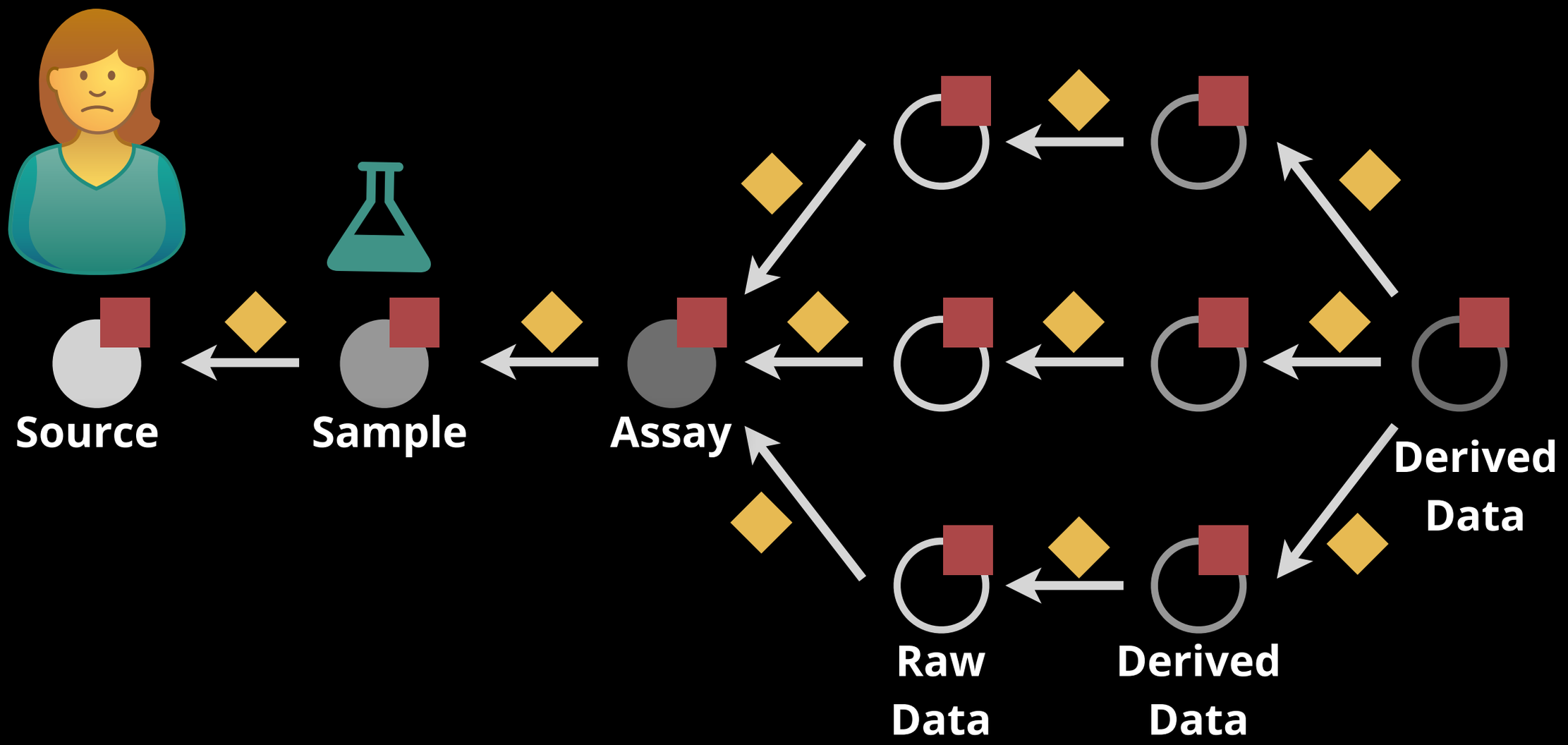
« 1 2 3 4 5 »

Workflow Engine

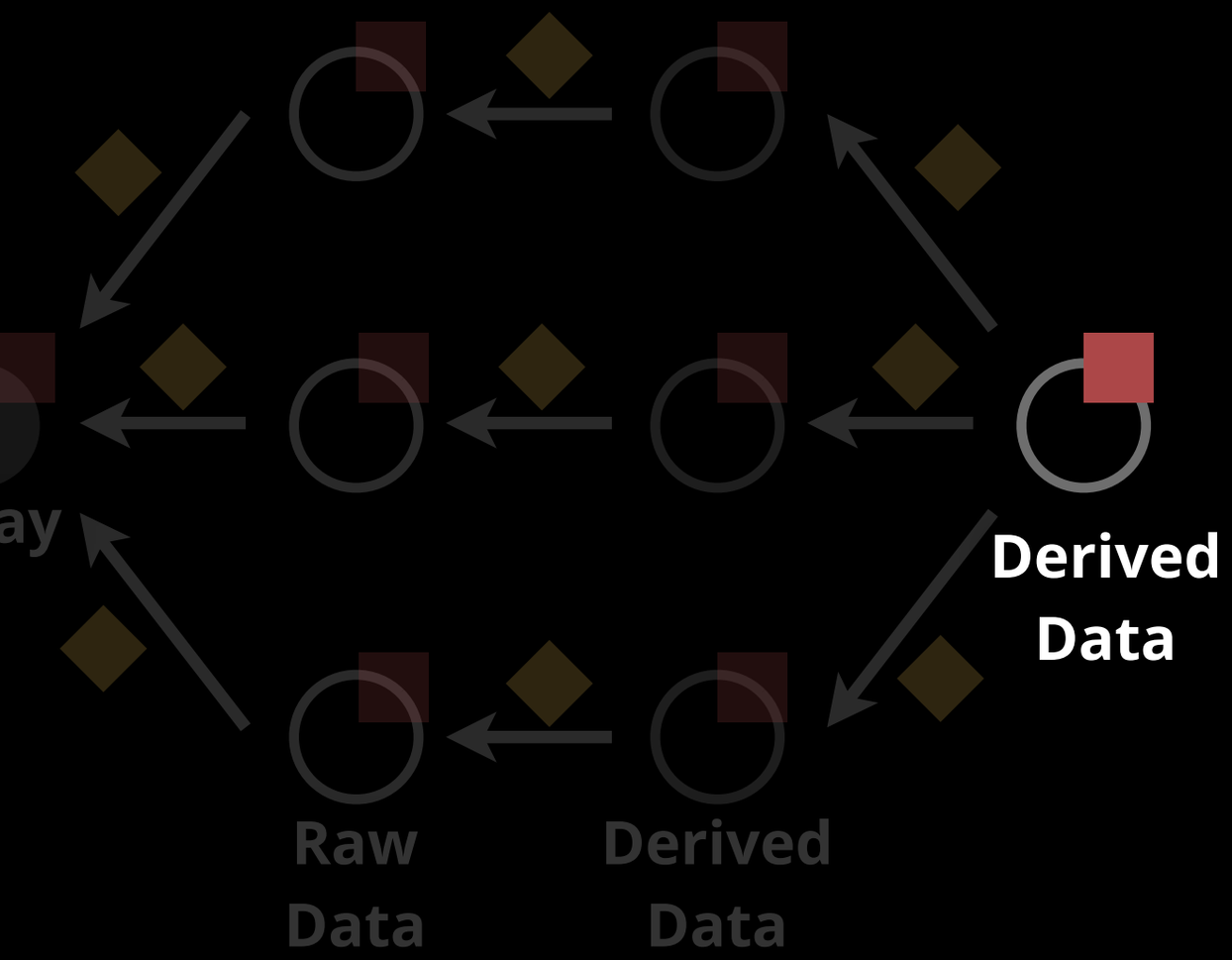
Provenance Tracking



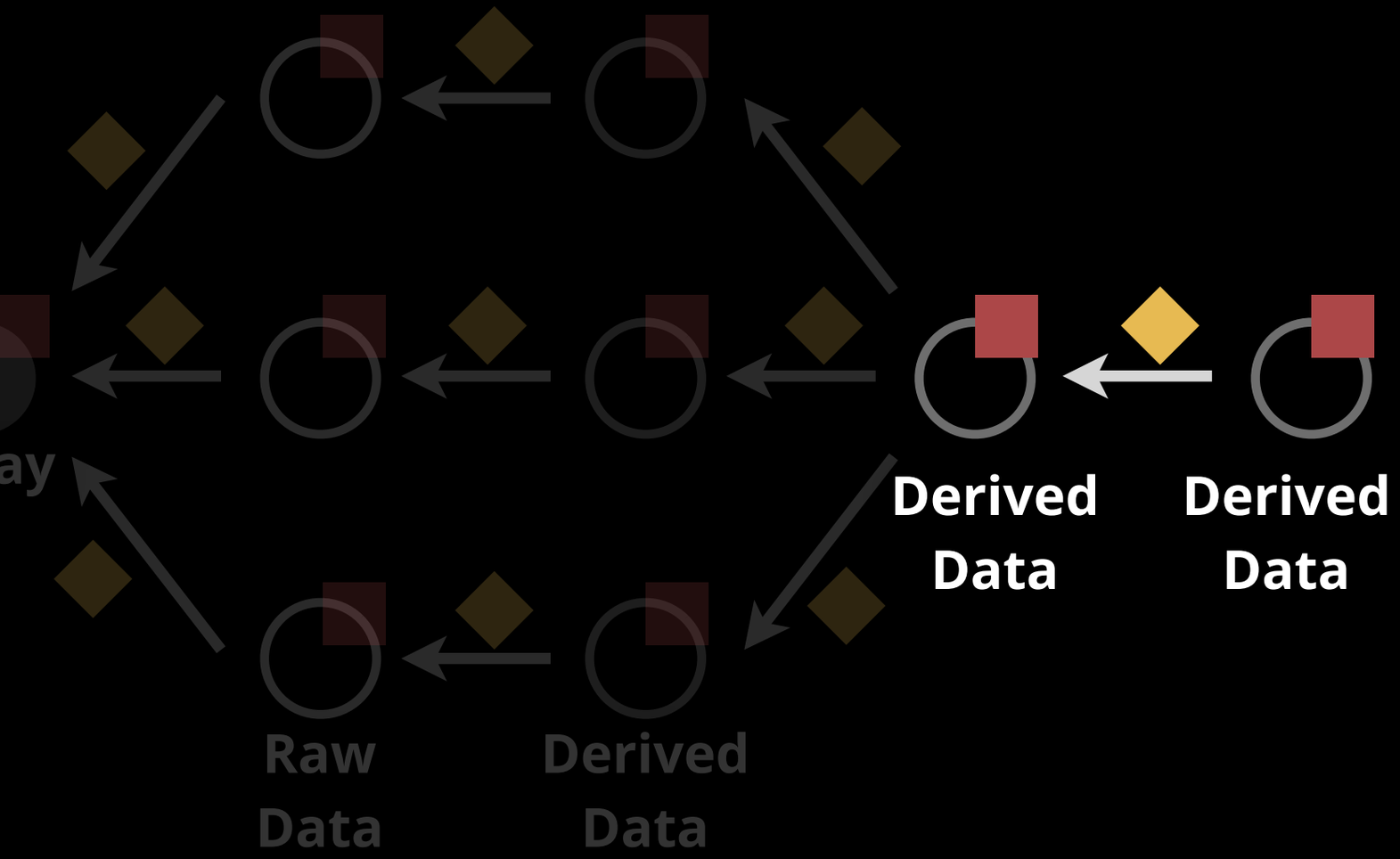
Workflow Engine



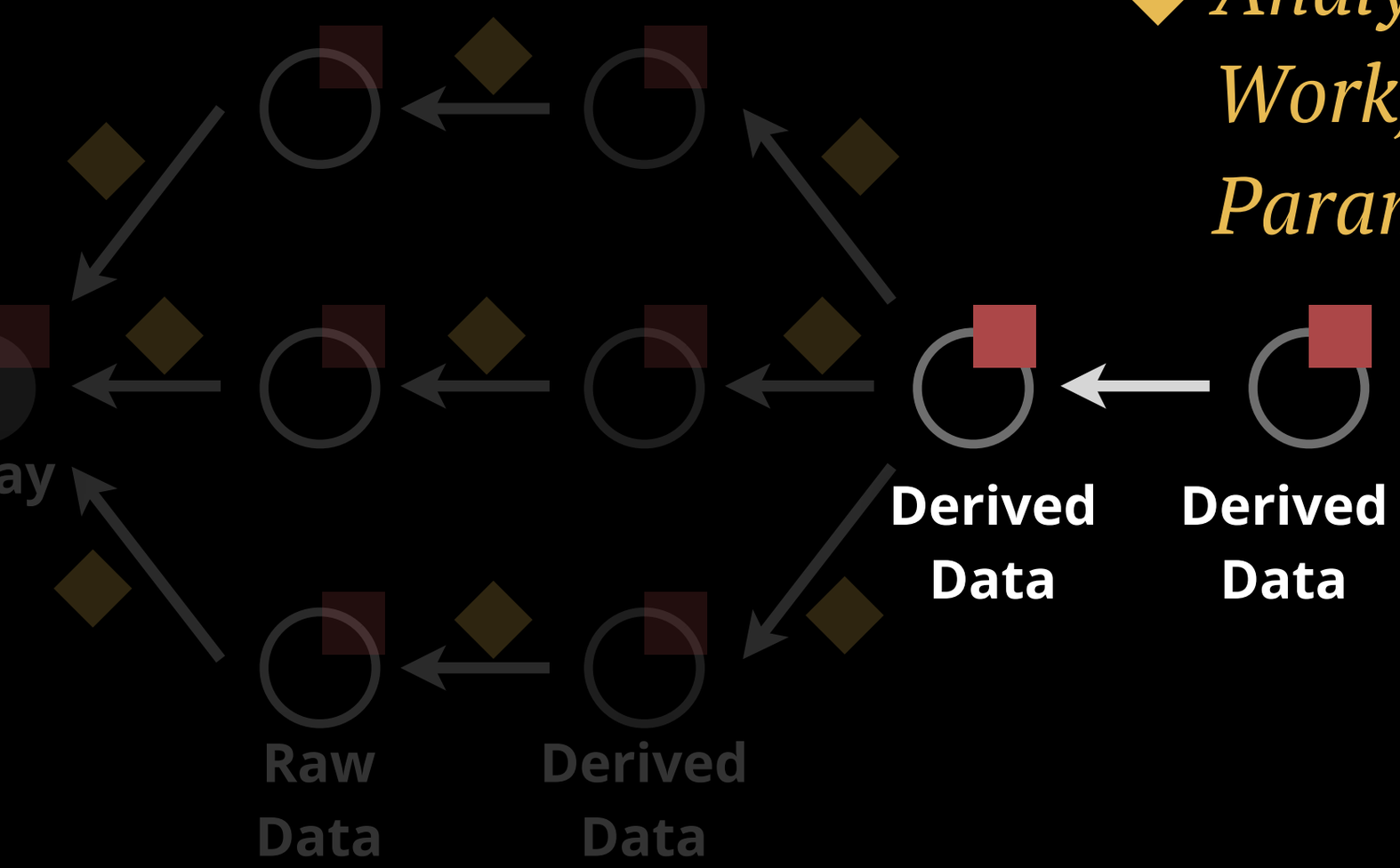
Workflow Engine



Workflow Engine



Workflow Engine

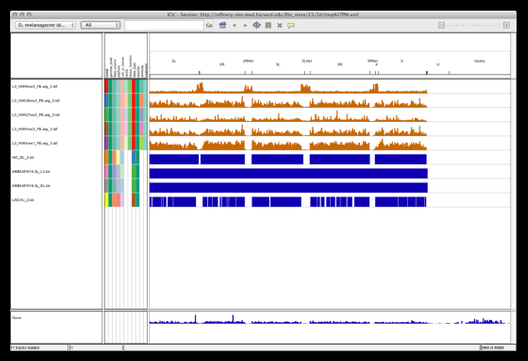


- *Analysis Details*
- *Workflow*
- *Parameters*

Visualization

file-driven

database-driven



Java Webstart tools *or*
built-in Refinery tools

built-in Refinery tools

IGV – Session: http://refinery-dev.med.harvard.edu/file_store/23/2d/tmpAll7PM.xml

D. melanogaster (d... | All | Go

NAME genome_build data_source platform cell_or_tissue factor factor_function data_type species antibody treatment

2L 2R 2RHet 3L 3LHet 3R 3RHet 4 X U Uextra

L3_H3K9me3_FE.wig_3.tdf
L3_H3K36me3_FE.wig_3.tdf
L3_H3K27me3_FE.wig_3.tdf
L3_H3K4me3_FE.wig_3.tdf
L3_H3K4me1_FE.wig_3.tdf
HiC_EL_2.bb
iHMM.M1K16.fly_L3.bb
iHMM.M1K16.fly_EL.bb
LAD.Kc_2.bb

Gene

11 tracks loaded | :1 | 94M of 466M

Refinery – Data Set ENCODE-X Chromatin

refinery-dev.med.harvard.edu/data_sets/chromatin/#

Reader

Refinery Home Statistics About Register

Data Set ENCODE-X Chromatin

Samples | Details | Sharing

5 of **1298** selected | View in IGV | Download as Archive | Showing Data Annotation

Facets | List | Matrix

Reset All

Columns | 10 20 50 100

Species

- D. melanogaster 5

Cell Or Tissue

- Third instart lavae (L3) 5

Factor Function

Factor

- H3K27me3 1
- H3K36me3 1
- H3K4me1 1
- H3K4me3 1
- H3K9me3 1

Antibody

Treatment

Data Type

Platform

- Illumina sequencing 5

Data Source

Genome Build

Species	Cell Or Tissue	Factor Function	Factor	Antibody	Treatment	Data Type	Platform	Data Source	Genome Build
D. melanogaster	Third instart lavae (L3)	Histone modification	H3K9me3	H3K9me3	None	Histone ChIP-seq	Illumina sequencing	modENCODE	dm3
D. melanogaster	Third instart lavae (L3)	Histone modification	H3K36me3	H3K36me3	None	Histone ChIP-seq	Illumina sequencing	modENCODE	dm3
D. melanogaster	Third instart lavae (L3)	Histone modification	H3K27me3	H3K27me3	None	Histone ChIP-seq	Illumina sequencing	modENCODE	dm3
D. melanogaster	Third instart lavae (L3)	Histone modification	H3K4me3	H3K4me3	None	Histone ChIP-seq	Illumina sequencing	modENCODE	dm3
D. melanogaster	Third instart lavae (L3)	Histone modification	H3K4me1	H3K4me1	None	Histone ChIP-seq	Illumina sequencing	modENCODE	dm3

« 1 2 3 4 5 »

IGV – Session: http://refinery-dev.med.harvard.edu/file_store/23/2d/tmpAll7PM.xml

D. melanogaster (d... | All | Go

NAME genome_build data_source platform cell_or_tissue factor factor_function data_type species antibody treatment

2L 2R 2RHet 3L 3LHet 3R 3RHet 4 X U Uextra

L3_H3K9me3_FE.wig_3.tdf
L3_H3K36me3_FE.wig_3.tdf
L3_H3K27me3_FE.wig_3.tdf
L3_H3K4me3_FE.wig_3.tdf
L3_H3K4me1_FE.wig_3.tdf
HiC_EL_2.bb
iHMM.M1K16.fly_L3.bb
iHMM.M1K16.fly_EL.bb
LAD.Kc_2.bb

Gene

11 tracks loaded | :1 | 94M of 466M

IGV – Session: http://refinery-dev.med.harvard.edu/file_store/23/2d/tmpAll7PM.xml

D. melanogaster (d... | chr2L | chr2L:22,430,000-22,450,000 | Go

NAME
genome_build
data_source
platform
cell_or_tissue
factor
factor_function
data_type
species
antibody
treatment

19 kb

22,432 kb | 22,434 kb | 22,436 kb | 22,438 kb | 22,440 kb | 22,442 kb | 22,444 kb | 22,446 kb | 22,448 kb

L3_H3K9me3_FE.wig_3.tdf [0 - 7.47]
L3_H3K36me3_FE.wig_3.tdf [0 - 4.14]
L3_H3K27me3_FE.wig_3.tdf [0 - 1.41]
L3_H3K4me3_FE.wig_3.tdf [0 - 9.15]
L3_H3K4me1_FE.wig_3.tdf [0 - 2.38]
HiC_EL_2.bb HP1_centromeric
iHMM.M1K16.fly_L3.bb 12_Het1 | 1_Pro | 8_Egn5 | 12_Het1 | 16_Low3 | 1_Pro | 12_Het1 | 16_Low3 | 16_Low3 | 16_Low3
iHMM.M1K16.fly_EL.bb 12_Het1 | 1_Pro | 9_Egn6 | 13_Het2 | 12_Het1 | 16_Low3 | 3_Enh2 | 9_Egn6 | 12_Het1 | 17_Unmap | 17_Unmap
LAD.Kc_2.bb

Gene CG17490 CG17493

11 tracks loaded | chr2L:22,440 | 135M of 466M

Refinery – Data Set ENCODE-X Chromatin

refinery-dev.med.harvard.edu/data_sets/8c6d6f16-4bba-11e2-9799-005056a07c79/#

Refinery | Home | Statistics | About | modENCODE

Data Set ENCODE-X Chromatin

Samples | Attributes | Analyses | Details | Sharing

12 of 12 selected | View in IGV | Download as Archive | Showing | Data | Annotation

Facets | List | Matrix

Reset All | Columns | 10 20 50 100

Species
Cell Or Tissue
Factor Function
Factor
Antibody
Treatment
Data Type
Platform
Data Source
Genome Build

Species	Cell Or Tissue	Factor Function	Factor	Antibody	Treatment	Data Type	Platform	Data Source	Genome Build
C. elegans	Larvae stage 3 (L3)					Chromatin state map	hiHMM	modENCODE	WS220
C. elegans	Early embryo (EE)					Chromatin state map	hiHMM	modENCODE	WS220
C. elegans	MXEMB					Lamina associated domains	LAD	modENCODE	WS220
D. melanogaster	Late embryo 14-16hr (EL)					HiC topological domains	HiC	Others	dm3
D. melanogaster	Third instar larvae (L3)					Chromatin state map	hiHMM	modENODE	dm3
D. melanogaster	Late embryo 14-16hr (EL)					Chromatin state map	hiHMM	modENCODE	dm3
D. melanogaster	Kc					Lamina associated domains	LAD	Others	dm3
H. sapiens	H1-hESC					HiC topological domains	HiC	Others	hg19
H. sapiens	H1-hESC					Chromatin state map	hiHMM	ENCODE	hg19
H. sapiens	GM12878					Chromatin state map	hiHMM	ENCODE	hg19

« 1 2 3 4 5 »

Refinery – Data Set ENCODE-X Chromatin

refinery-dev.med.harvard.edu/data_sets/chromatin/#

Reader

Refinery Home Statistics About Register

Data Set ENCODE-X Chromatin

Samples | Details | Sharing

13 of **1298** selected | View in IGV | Download as Archive | Showing Data Annotation

Facets | List | Matrix

Reset All

Columns | 10 20 50 100

Species
- H. sapiens 8
- D. melanogaster 5

Cell Or Tissue
- GM12878 8
- Third instar lavae (L3) 5

Factor Function

Factor
- H3K27me3 3
- H3K36me3 3
- H3K4me3 3
- H3K4me1 2
- H3K9me3 2

Antibody

Treatment

Data Type

Platform
- Illumina sequencing 13

Data Source

Species	Cell Or Tissue	Factor Function	Factor	Antibody	Treatment	Data Type	Platform	Data Source	Genome Build
D. melanogaster	Third instar lavae (L3)	Histone modification	H3K9me3	H3K9me3	None	Histone ChIP-seq	Illumina sequencing	modENCODE	dm3
D. melanogaster	Third instar lavae (L3)	Histone modification	H3K36me3	H3K36me3	None	Histone ChIP-seq	Illumina sequencing	modENCODE	dm3
D. melanogaster	Third instar lavae (L3)	Histone modification	H3K27me3	H3K27me3	None	Histone ChIP-seq	Illumina sequencing	modENCODE	dm3
D. melanogaster	Third instar lavae (L3)	Histone modification	H3K4me3	H3K4me3	None	Histone ChIP-seq	Illumina sequencing	modENCODE	dm3
D. melanogaster	Third instar lavae (L3)	Histone modification	H3K4me1	H3K4me1	None	Histone ChIP-seq	Illumina sequencing	modENCODE	dm3
H. sapiens	GM12878	Histone modification	H3K36me3	H3K36me3	None	Histone ChIP-seq	Illumina sequencing	ENCODE	hg19
H. sapiens	GM12878	Histone modification	H3K27me3	H3K27me3	None	Histone ChIP-seq	Illumina sequencing	ENCODE	hg19
H. sapiens	GM12878	Histone modification	H3K4me3	H3K4me3	None	Histone ChIP-seq	Illumina sequencing	ENCODE	hg19
H. sapiens	GM12878	Histone modification	H3K4me1	H3K4me1	None	Histone ChIP-seq	Illumina sequencing	ENCODE	hg19
H. sapiens	GM12878	Histone modification	H3K36me3	H3K36me3	None	Histone ChIP-seq	Illumina sequencing	ENCODE	hg19
H. sapiens	GM12878	Histone modification	H3K27me3	H3K27me3	None	Histone ChIP-seq	Illumina sequencing	ENCODE	hg19
H. sapiens	GM12878	Histone modification	H3K4me3	H3K4me3	None	Histone ChIP-seq	Illumina sequencing	ENCODE	hg19
H. sapiens	GM12878	Histone modification	H3K9me3	H3K9me3	None	Histone ChIP-seq	Illumina sequencing	ENCODE	hg19

« 1 2 3 4 5 »

Refinery – Data Set ENCODE-X Chromatin

refinery-dev.med.harvard.edu/data_sets/chromatin/#

Reader

Refinery Home Statistics About Register

Data Set ENCODE-X Chromatin

Samples Details Sharing

13 of 1298 selected View in IGV Download as Archive Showing Data Annotation

Facets List Matrix

Reset All Columns 10 20 50 10

Species
- H. sapiens 8
- D. melanogaster 5

Cell Or Tissue
- GM12878 8
- Third instart lavae (L3) 5

Factor Function

Factor
- H3K27me3 3
- H3K36me3 3
- H3K4me3 3
- H3K4me1 2
- H3K9me3 2

Antibody

Treatment

Data Type

Platform
- Illumina sequencing 13

Data Source

Species	Cell						Platform	Data Source	Genome Build
D. melanogaster	Third					ChIP-seq	Illumina sequencing	modENCODE	dm3
D. melanogaster	Third					ChIP-seq	Illumina sequencing	modENCODE	dm3
D. melanogaster	Third					ChIP-seq	Illumina sequencing	modENCODE	dm3
D. melanogaster	Third instart lavae (L3)	Histone modification	H3K4me1	H3K4me1	None	Histone ChIP-seq	Illumina sequencing	modENCODE	dm3
H. sapiens	GM12878	Histone modification	H3K36me3	H3K36me3	None	Histone ChIP-seq	Illumina sequencing	ENCODE	hg19
H. sapiens	GM12878	Histone modification	H3K27me3	H3K27me3	None	Histone ChIP-seq	Illumina sequencing	ENCODE	hg19
H. sapiens	GM12878	Histone modification	H3K4me3	H3K4me3	None	Histone ChIP-seq	Illumina sequencing	ENCODE	hg19
H. sapiens	GM12878	Histone modification	H3K4me1	H3K4me1	None	Histone ChIP-seq	Illumina sequencing	ENCODE	hg19
H. sapiens	GM12878	Histone modification	H3K36me3	H3K36me3	None	Histone ChIP-seq	Illumina sequencing	ENCODE	hg19
H. sapiens	GM12878	Histone modification	H3K27me3	H3K27me3	None	Histone ChIP-seq	Illumina sequencing	ENCODE	hg19
H. sapiens	GM12878	Histone modification	H3K4me3	H3K4me3	None	Histone ChIP-seq	Illumina sequencing	ENCODE	hg19
H. sapiens	GM12878	Histone modification	H3K9me3	H3K9me3	None	Histone ChIP-seq	Illumina sequencing	ENCODE	hg19

« 1 2 3 4 5 »

Launch IGV

You selected samples from 2 different genome builds. To view the samples, open IGV with the corresponding genome.

H. sapiens (hg19) D. melanogaster (dm3)

Close

IGV – Session: http://refinery-dev.med.harvard.edu/file_store/b1/e0/tmpd1ngRP.xml

Human hg19 | chr12 | chr12:32,110,000-32,140,000 | Go

p13.32 p13.2 p12.3 p12.1 p11.21 q12 q13.11 q13.2 q14.2 q15 q21.2 q21.32 q22 q23.2 q24.11 q24.22 q24.32

29 kb

0 kb | 32,120 kb | 32,130 kb | 32,1

NAME genome_build data_source platform cell_or_tissue factor factor_function data_type species antibody treatment

GM12878_H3K36me3_Broad_None.wig_3.tdf [0 - 12]

GM12878_H3K27me3_Broad_None.wig_3.tdf [0 - 1.76]

GM12878_H3K4me3_Broad_None.wig_3.tdf [0 - 21]

GM12878_H3K4me1_Broad_None.wig_3.tdf [0 - 7.00]

GM12878_H3K36me3_UW_None.wig_3.tdf [0 - 10]

GM12878_H3K27me3_UW_None.wig_3.tdf [0 - 2.71]

GM12878_H3K4me3_UW_None.wig_3.tdf [0 - 69]

GM12878_H3K9me3_Broad_None.wig_3.tdf [0 - 16]

HiC_hESC_2.bb

iHMM.M1K16.human_H1.bb

_Low2 1_Pro 4_Egn1 17_Unmap 17_Unmap 4_Egn1 4_Egn1 4_Egn1 7_Egn4 4_Egn1 4_Egn1 17_Unmap 9_Egn6 12_Het1 5_Egn2

iHMM.M1K16.human_GM.bb

Het1 1_Pro 4_Egn1 17_Unmap 4_Egn1 4_Egn1 17_Unmap 12_Het1 12_Het1 12_Het1 12_Het1 12_Het1 2_En

HiC_IMR90_2.bb

LAD_fibroblast_2.bb

RefSeq Genes

KIAA1551

15 tracks loaded

94M of 450M

IGV – Session: http://refinery-dev.med

Human hg19 | chr12 | chr12:32,110,000-32,140,000 | Go

p13.32 p13.2 p12.3 p12.1 p11.21

0 kb

NAME genome_build data_source platform cell_or_tissue factor factor_function data_type species antibody treatment

GM12878_H3K36me3_Broad_None.wig_3.tdf [0 - 12]
GM12878_H3K27me3_Broad_None.wig_3.tdf [0 - 1.76]
GM12878_H3K4me3_Broad_None.wig_3.tdf [0 - 21]
GM12878_H3K4me1_Broad_None.wig_3.tdf [0 - 7.00]
GM12878_H3K36me3_UW_None.wig_3.tdf [0 - 10]
GM12878_H3K27me3_UW_None.wig_3.tdf [0 - 2.71]
GM12878_H3K4me3_UW_None.wig_3.tdf [0 - 69]
GM12878_H3K9me3_Broad_None.wig_3.tdf [0 - 16]
HiC_hESC_2.bb
iHMM.M1K16.human_H1.bb _Low2 1_Pro 4_Egn1 17_Unmap 17_Unmap
iHMM.M1K16.human_GM.bb Het1 1_Pro 4_Egn1 17_Unmap
HiC_IMR90_2.bb
LAD_fibroblast_2.bb

RefSeq Genes

15 tracks loaded

IGV – Session: http://refinery-dev.med.harvard.e

D. melanogaster (d... | chr2L | chr2L:22,430,000-22,450,000 | Go

0 kb 22,432 kb 22,434 kb 22,436 kb

NAME genome_build data_source platform cell_or_tissue factor factor_function data_type species antibody treatment

L3_H3K27me3_FE.wig_3.tdf [0 - 1.41]
L3_H3K36me3_FE.wig_3.tdf [0 - 4.14]
L3_H3K4me1_FE.wig_3.tdf [0 - 2.38]
L3_H3K4me3_FE.wig_3.tdf [0 - 9.15]
L3_H3K9me3_FE.wig_3.tdf [0 - 7.47]
HiC_EL_2.bb
iHMM.M1K16.fly_L3.bb 12_Het1 1_Pro 8_Egn5
iHMM.M1K16.fly_EL.bb 12_Het1 1_Pro 9_Egn6 13_Het2
LAD.Kc_2.bb

Gene CG17490

11 tracks loaded | chr2L:22,431,312 | 94M of 450M

REFINERY PLATFORM

http://www.refinery-platform.org
http://www.github.com/parklab/refinery-platform

Work in progress!
First release scheduled for late summer

Thank you open source developers for your code!

PP57 | Visual Exploration for Cancer Subtype Analysis (w/ StratomeX)
Tuesday, 10:30 am - Hall 7

WK6 | Integrated Visualization and Computing for Biomedical Research
Tuesday, 2:10 pm - Roof Garden (w/ Jeremy Goecks & Dave Clements)