Accessing modENCODE data Data Coordination Centre Tools

http://www.modencode.org help@modencode.org

earch modENCODE Data Sets

1 Department of Genetics, University of Cambridge, Cambridge, UK 2 Lawrence Berkeley National Laboratory; Genomics Division, Berkeley, CA, USA 3 Ontario Institute for Cancer Research, Toronto, ON, Canada

> Links to summary of all experiments and data

> > generated..

Genomic Region Search

Explore an hierarchical view of the

Analyze Lists of Data

physical regulatory networks.

relationshops in our analysis

Description of the submission

and its location in the context

of the modENCODE project.

External Links

Single click queries find

overlapping and nearby genes for features generated by the

submission.

Data Files Download

widgets. See an example.

Explore and

lists of

discover

Features can be

exported in various

Analyze, Upload

identifiers to use

in queries and

genomic region for

Park Lab

Includes folded view and also data

Use Template

queries using our predefined

templates have been designed

around common tasks performed

To see how they work, why not try

Questions? Comments? Click here!

Powered by

searches. These customizable

by our biologist community.

a template from our examples

View GBrowse tracks and get

raw data files from the FTP

site.

about DHS, TSS, replication, etc.

features found by the

Genomic Region Search

The modENCODE project aims to identify all sequence-based functional elements in the C. elegans and D.

experimental details, download results and access the GBrowse genome browser. Explore some of the tools

modMine is an integrated web resource of data & tools to browse and search modENCODE data and

Fly Chromatin states

melanogaster genomes. modENCODE labs submit data to the Data Coordination Center (DCC) where we organize

modMine > Home

Genome Region Search:

find features in user

assigned genomic

regions

Interactive exploratory

interface of the regulatory

regions

Sergio Contrino¹, Jelena Aleksic¹, Daniela Butano¹, Fengyuan Hu¹, Alex Kalderimis¹, Ellen Kephart², Paul Lloyd², Rachel Lyne¹, Marc Perry³, Peter Ruzanov³, Richard Smith¹, Radek Stepan¹, Julie Sullivan¹, Quang Trinh³, Nicole Washington², Suzanna Lewis², Gos Micklem¹, Lincoln Stein³.

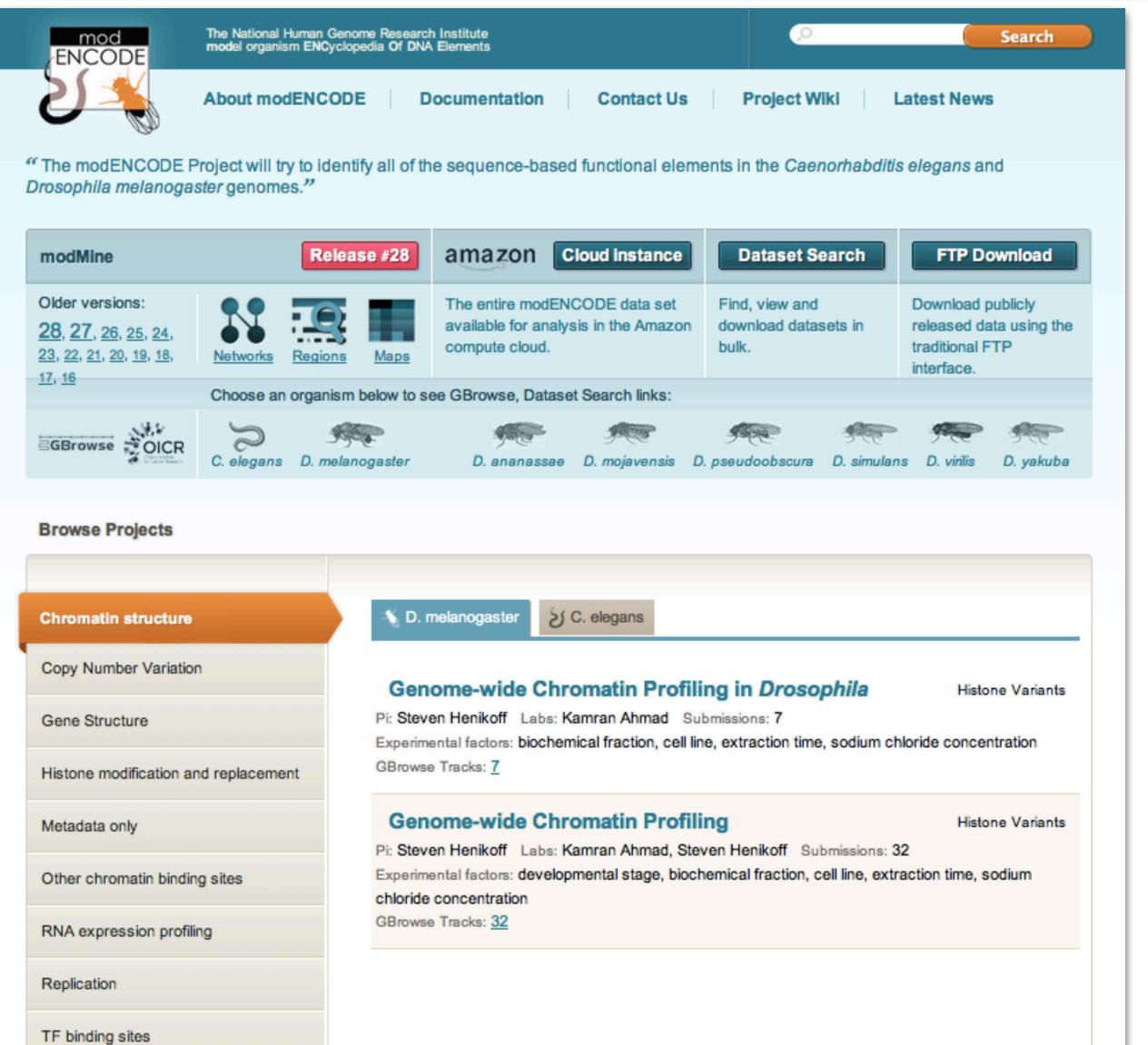
The DCC is...

The Data Coordination Centre vets the raw and interpreted data produced by the modENCODE project to ensure consistency and completeness. The data is then released to the community via the archival databases Gene Expression Omnibus (GEO) and the Short Read Archive (SRA), as well as the model organism databases WormBase and FlyBase.

The DCC also developed various tools for accessing the data produced by the project, and this poster describes how the community can use three of these tools: the modMine integrated data portal, a GBrowse genome browser, and a faceted submission filtering tool. A fourth tool is a specialised Wiki. All are accessible through http://www.modencode.org.

A key feature of the data management of the project is the extensive and controlled collection of experimental metadata. The use of the metadata enables fine-grained control over presentation and queries, and allows end users of the data to find more easily and precisely what they need.

Commonly meta-data is captured in free text which makes it hard to use in queries and for presentation of results. The fact that the data orginates from a limited number of data producers has made it easier to collect meta-data in a very controlled fashion within the modENCODE project.



Dataset Search

Dataset Search: find, view and download datasets.

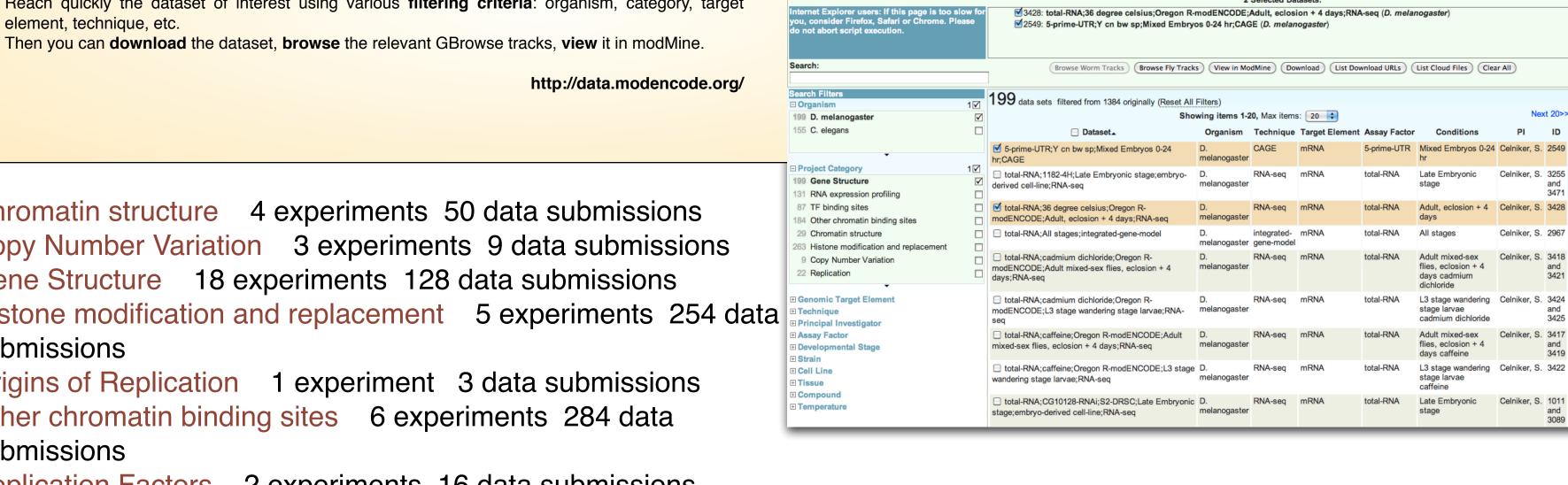
Reach quickly the dataset of interest using various filtering criteria: organism, category, target element, technique, etc.

Chromatin structure 4 experiments 50 data submissions Copy Number Variation 3 experiments 9 data submissions Gene Structure 18 experiments 128 data submissions Histone modification and replacement 5 experiments 254 data submissions

Origins of Replication 1 experiment 3 data submissions Other chromatin binding sites 6 experiments 284 data submissions

Replication Factors 2 experiments 16 data submissions Replication Timing 2 experiments 4 data submissions RNA expression profiling 26 experiments 682 data submissions

TF binding sites 3 experiments 390 data submissions



Contact Us

Welcome to modMine

Enter names, identifiers or keywords for genes, proteins, pathways, ontology terms, etc. (e.g.

eriments by type, lab

NA-seq, Snyder, CP190)

and present the results.

provided below.

Expression heatmap for a list

of selected genes

Fly Gene Expression

View an expression score heatmap for any list of fly genes. See an

To upload your own list of genes, use the form above or go to the 'Lists' tab and click on 'Upload' to

create and name the new list. You

Upload Lists

CG9151, FBgn0000099,

CG3629, TfIIB, Mad, CG1775,

upload

submission view

Submission: seq-AB1791 H3 N2 L3

Experiment Type: ChIP-seq

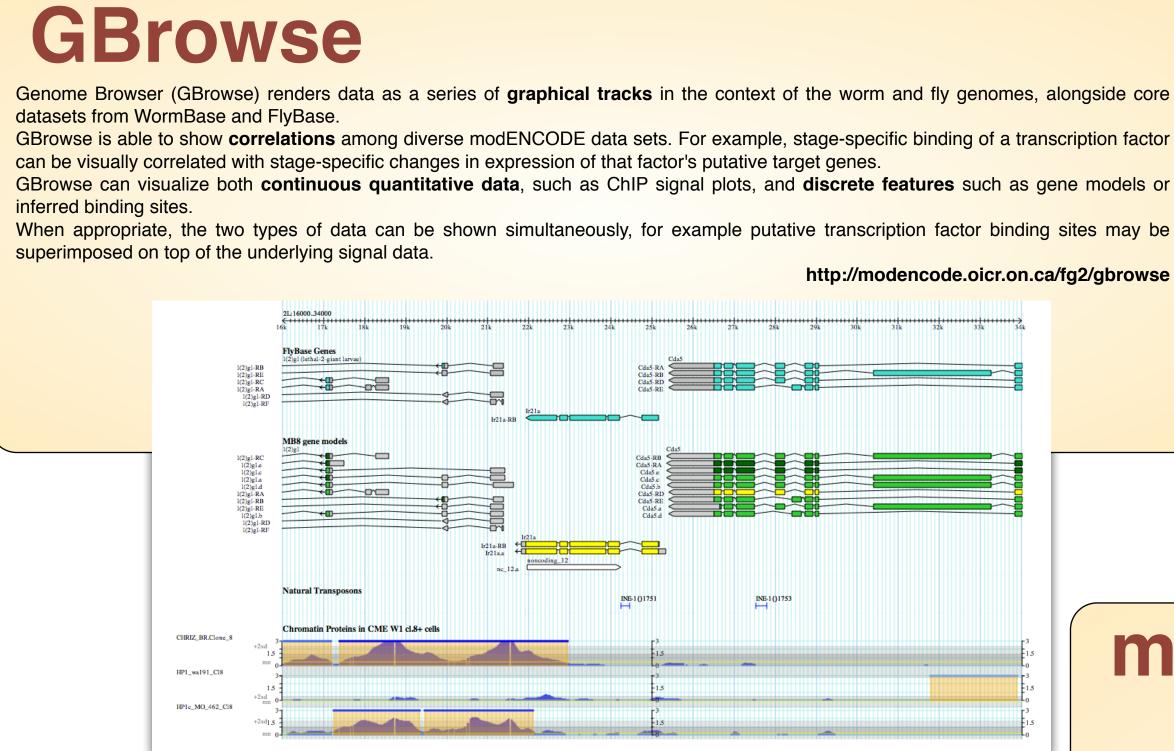
LIEB_WIG_CHIPSEQ_COREHIS LIEB_WIG_CHIPSEQ_COREHIS

Enter a list of identifiers

can also use any of modMine's

Keyword Search: ordered

faceted results



modMine

 is a data warehouse that integrates modENCODE results with data from FlyBase, WormBase and other sources. provides easy navigation through modENCODE experiments to view and export data that have been submitted and vetted in the

project. allows keywords searches on experimental details - includes a library of pre-defined template queries to mine modENCODE

allows you to create and analyse lists of data and upload your

is based on the InterMine data warehousing system.

modMine views

Summary information

about experiment

experiment view **Genomic Distributions of Histor**

two embryonic stages, and will then extend the analysis of a subset of proteins to four additional animal tissues/stages Export features from all BindingSite 761178 ○ TAB CSV GFF3 SEQ All 24 tracks \chi X 2L 2R 3L 3R 4 L submissions in GFF3, csv or All entries in public repositories generated by this experiment other formats Access data generated by each View and sort table by submission experimental factors antibodies, cell lines, View all submissions for this development stages, etc experiment - each submission varies Karpen_CHROMPROTS_BG3_Binding_Sites = experimental conditions 921 BEAF-70.BG3

We aim to determe the locations of 125 chromosomal proteins and histone modifications across the Drosophila melanogaster genome. The proteins and modifications under study are involved in basic chromosomal functions such

All the available data files 253.pirrota 435 B4 INPUT.CEL 255.pirrota 435 B4 INPUT.CEL 265.pirrota 439 B3 BEAF.CEL. 319.walues.wig @ 319.pirrota 498 B4 BEAF-70.c repset.4620429.smoothedx.clu can be reached with

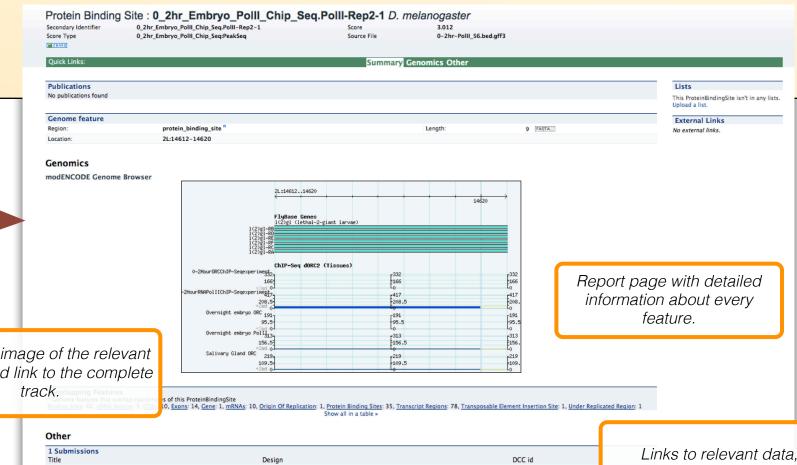
external links from the

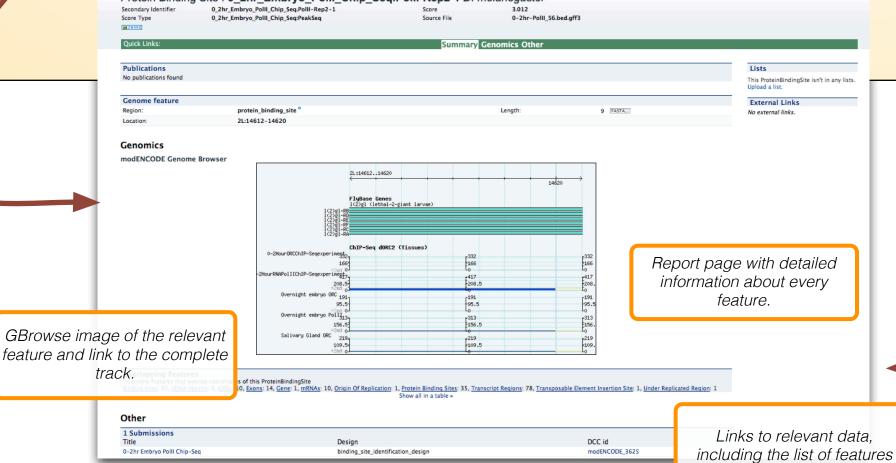
summary table.

overlapping the given feature

matin ImmunoPrecipitation (ChIP) using genomic tiling arrays. We will initially assay localizations using chromatin from three cell lines and

feature view





Web Service API

Templates query:

accessible via a simple web form to perform a specific task.

emplates are predefined queries, each has a simple form and a description. Click on a template to run it, you can search for templates by keyword and filter them

in to mark items as favourites 😭

Template queries can be

run on whole lists of data

by clicking on the

checkbox. List of the correct type will appear

automatically..

modifying an existing template. Or just ask us!

Search the library of

results are ordered by

relevance.

All Strains used in the modENCODE project

The table of results

allows quick export of

the data in several formats and the

creation of lists.

Bg3_Replication_Origins.MA2C_peak_1 46.16

Bg3_Replication_Origins.MA2C_peak_3 47.73

Bg3_Replication_Origins.MA2C_peak_4 719.35

Bg3_Replication_Origins.MA2C_peak_6 214.29

Bg3_Replication_Origins.MA2C_peak_7 225.45

Bg3_Replication_Origins.MA2C_peak_8 251.91

Bg3_Replication_Origins.MA2C_peak_9 49.5

Bg3_Replication_Origins.MA2C_peak_10 42.39

Bg3 Replication Origins.MA2C peak 11 210.84

Bg3_Replication_Origins.MA2C_peak_5 30.24

Use 'EXPORT' from the menu below to download the results of this query in a choice of formats.

modMine includes a RESTful web service API allowing you to access data directly from programs you write in Perl, Java, Python or Ruby. See http://intermine.org/wiki/WebServiceQuickStart

Library of Common Queries

modMine provides an extensive and easily extendable library of predefined ('canned') queries. Each query is

Create your own MyMine account to log in and build your own templates queries. An easy way to start is by





You can use '*' as a

wildcard when

entering text

constraint values.

ML-DmBG3-c2 ML-DmBG3-c2

ML-DmBG3-c2

ML-DmBG3-c2

ML-DmBG3-c2

ML-DmBG3-c2

ML-DmBG3-c2

ML-DmBG3-c2

Show Results Edit Query

NEW: Embed this query. Help You are not logged in. Log in to

modENCODE_711

modENCODE_711 modENCODE_711

modENCODE_711 cell line

modENCODE 711 cell line

modENCODE_711 cell line

modENCODE_711 cell line

modENCODE_711 cell line

...Interactive widgets

are used to summarise

properties of list of

features,

Click on the ∑ symbol in

the header and summary

information about the

relevant column will be

displayed.

2018758

2200539

