Accessing modENCODE data
Data Coordination Centre Tools

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

Sergio Contrino[1], Jelena Aleksic[1], Daniela Butano[1], Fengyuan Hu[1], Alex Kalderimis[1], Ellen Kephart[2], Paul Lloyd[2], Rachel Lyne[1], Marc Perry[3], Peter Ruzanov[3], Richard Smith[1], Radek Stepan[1], Julie Sullivan[1], Quang Trinh[3], Nicole Washington[2], Suzanna Lewis[2], Gos Micklem[1], Lincoln Stein[3].

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

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.
Then you can **download** the dataset, **browse** the relevant GBrowse tracks, **view** it in modMine.

http://data.modencode.org/
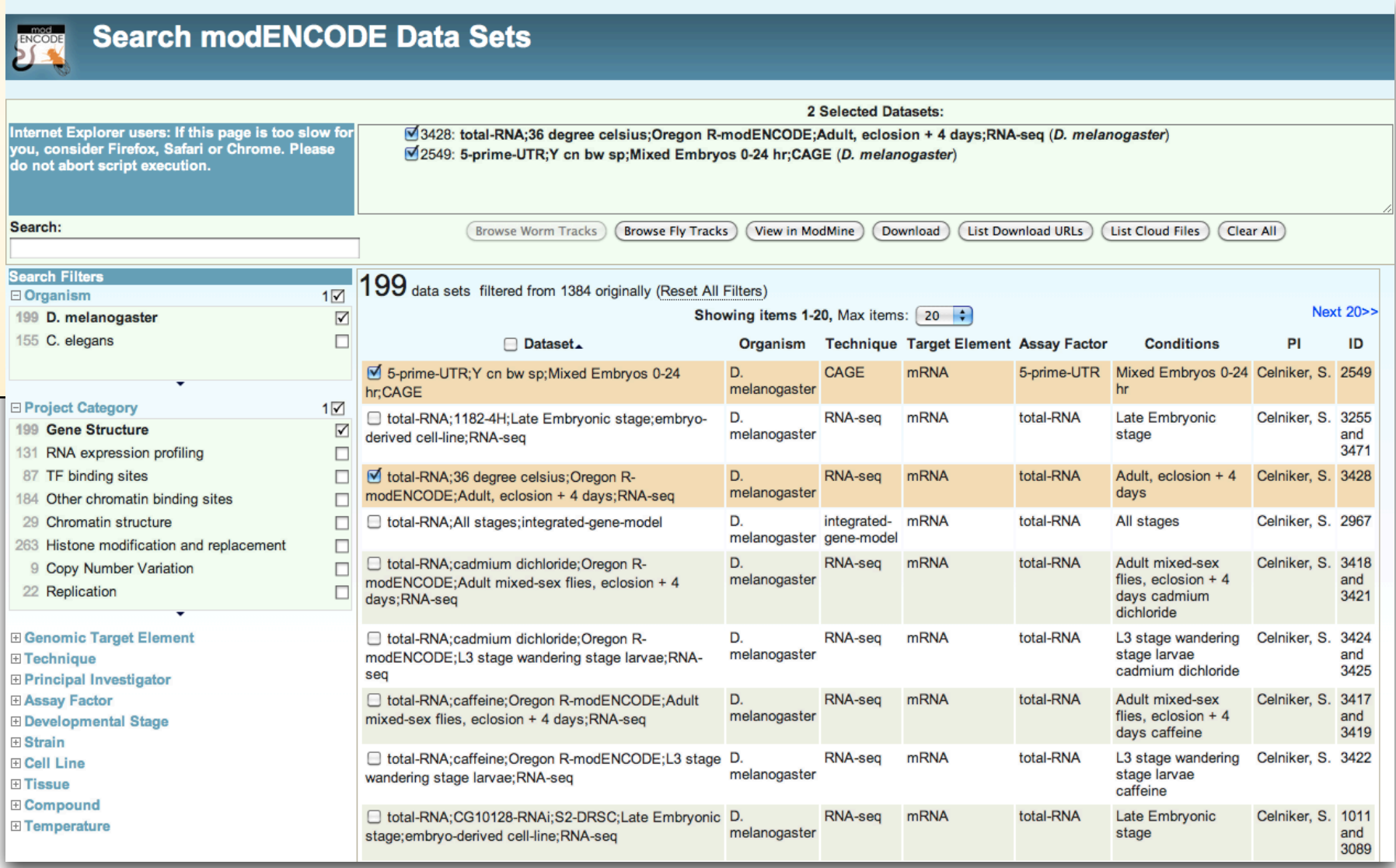
Chromatin structure   4 experiments 50 data submissions
Copy Number Variation   3 experiments 9 data submissions
Gene Structure   18 experiments 128 data submissions
Histone modification and replacement   5 experiments 254 data submissions
Origins of Replication   1 experiment 3 data submissions
Other 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

GBrowse

Genome Browser (GBrowse) renders data as a series of **graphical tracks** in the context of the worm and fly genomes, alongside core datasets from WormBase and FlyBase.
GBrowse is able to show **correlations** among diverse modENCODE data sets. For example, stage-specific binding of a transcription factor can be visually correlated with stage-specific changes in expression of that factor's putative target genes.
GBrowse can visualize both **continuous quantitative data**, such as ChIP signal plots, and **discrete features** such as gene models or inferred binding sites.
When appropriate, the two types of data can be shown simultaneously, for example putative transcription factor binding sites may be superimposed on top of the underlying signal data.

http://modencode.oicr.on.ca/fg2/gbrowse
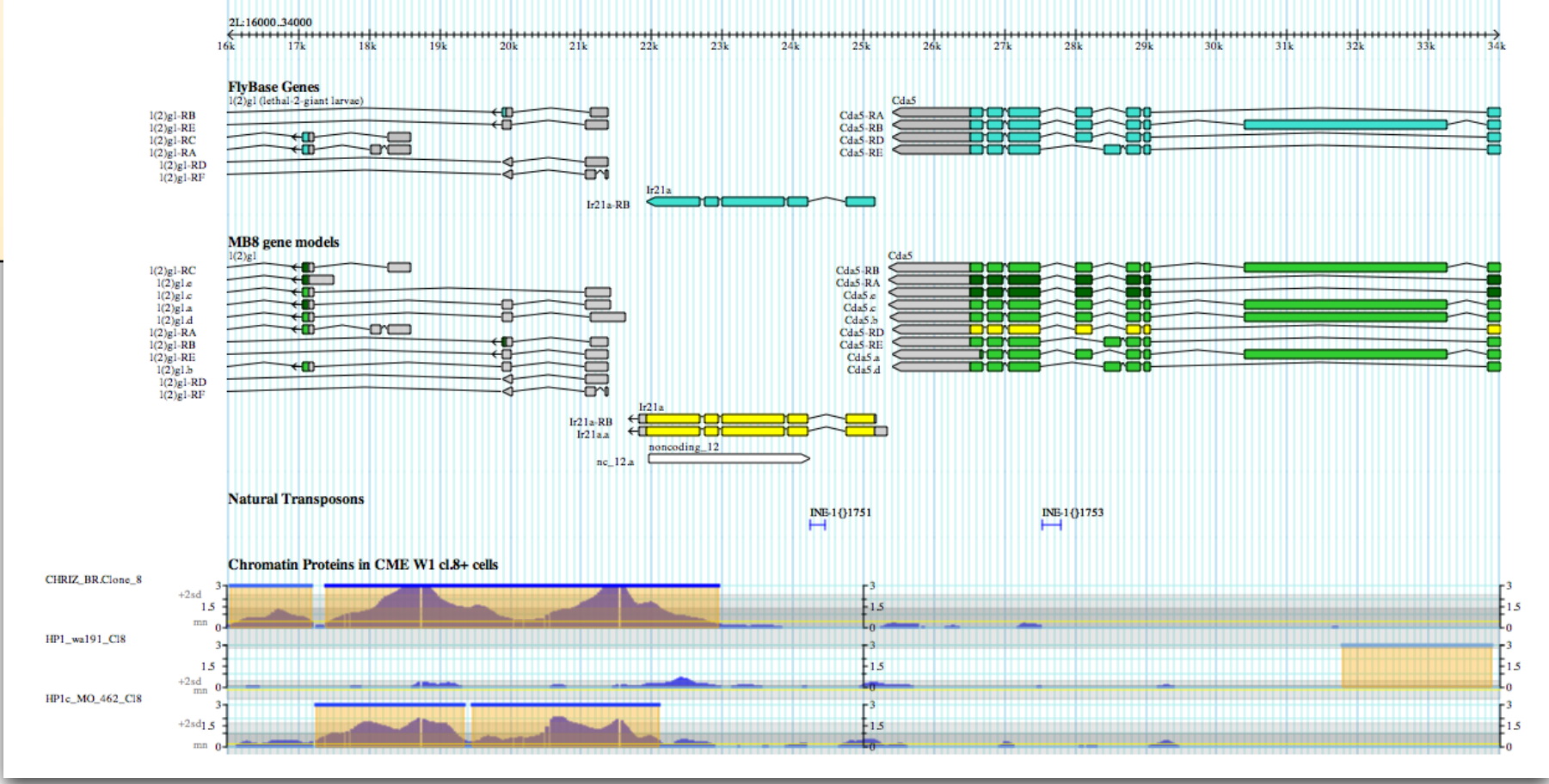
modMine

- 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 own lists
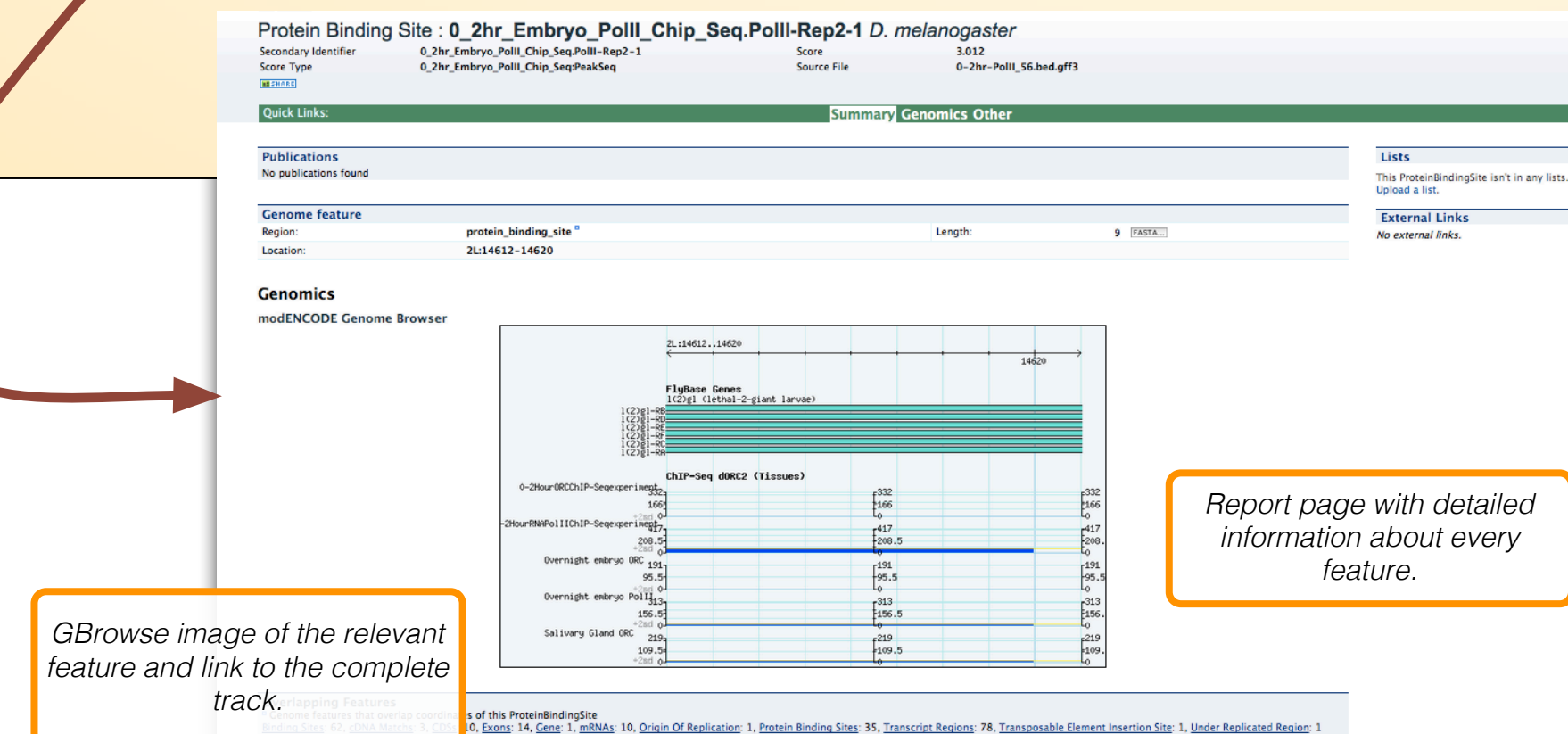- is based on the **InterMine** data warehousing system.
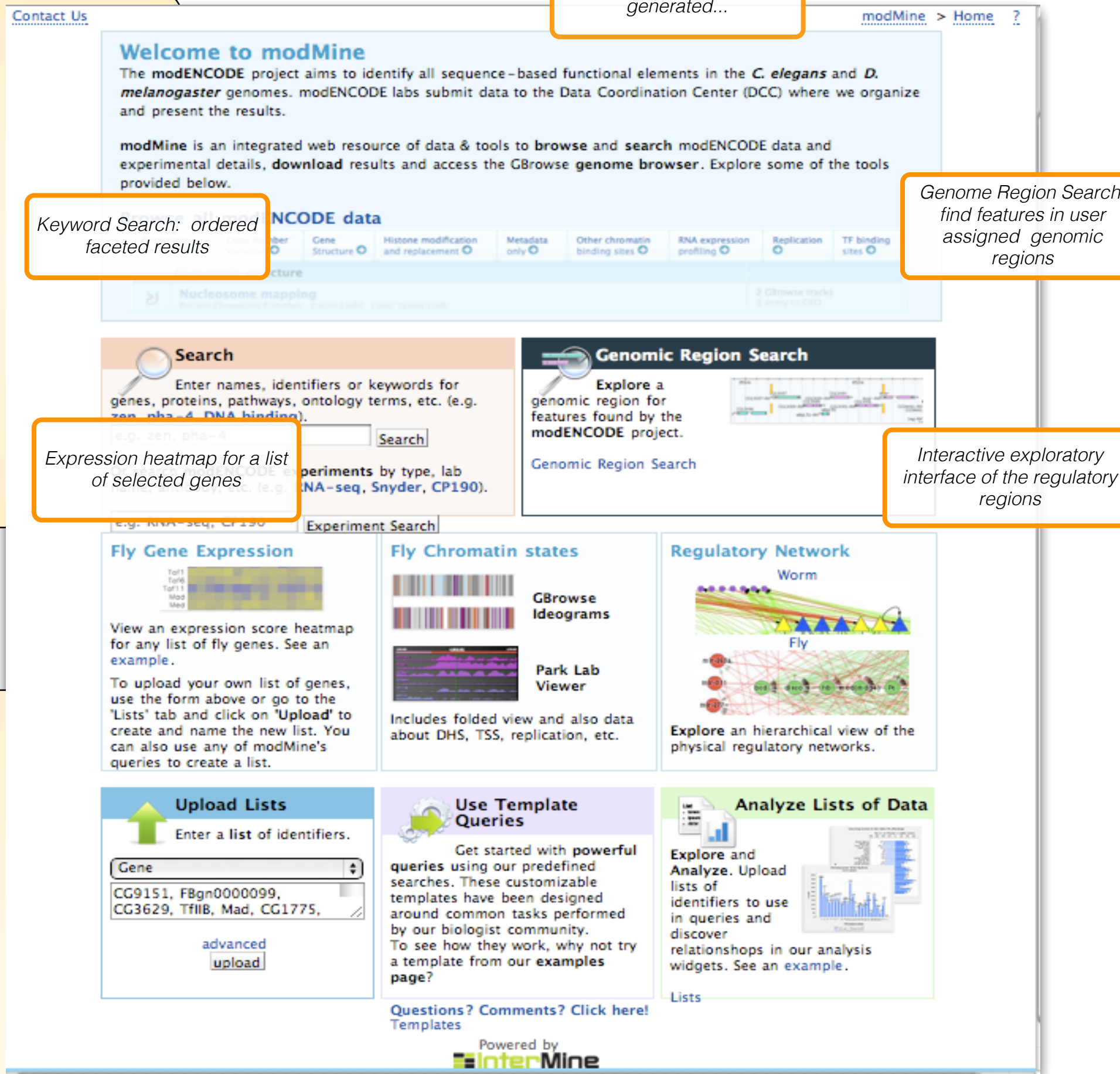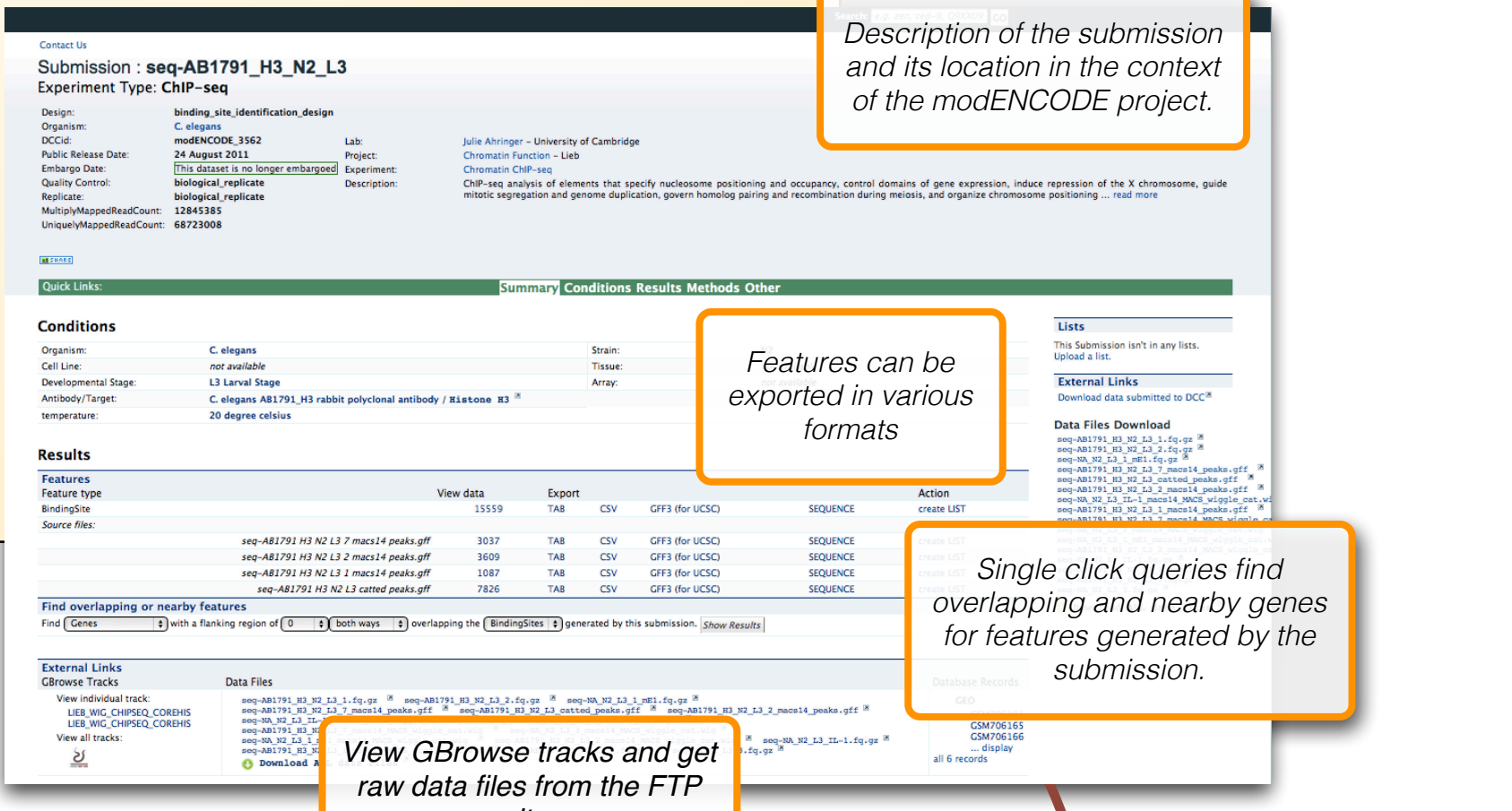

modMine views
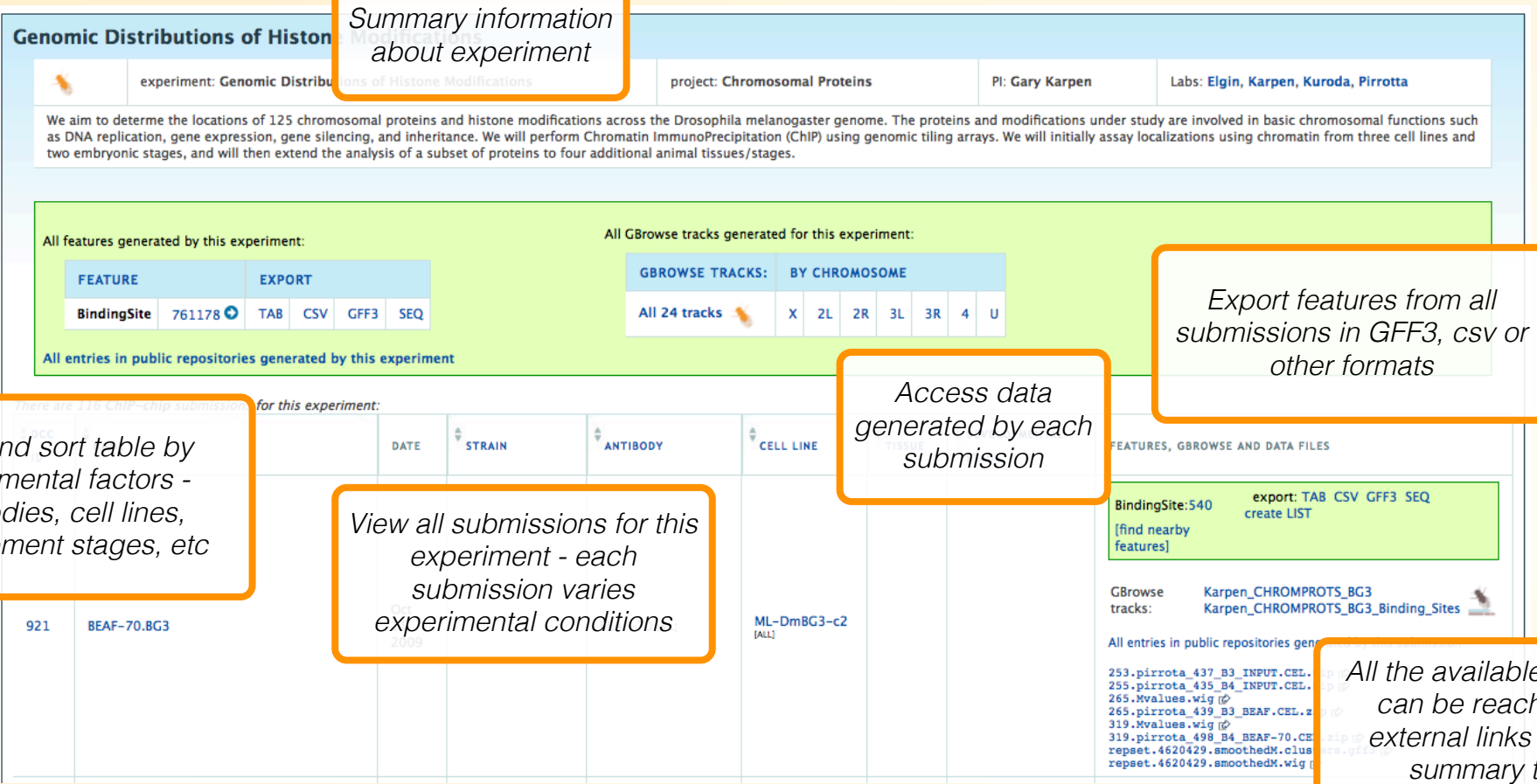
experiment view

submission view

feature view

Library of Common Queries

Templates query:
modMine provides an extensive and easily extendable library of predefined ('canned') queries. Each query is accessible via a simple web form to perform a specific task.
Create your own MyMine account to log in and build your own templates queries. An easy way to start is by modifying an existing template. Or just ask us!
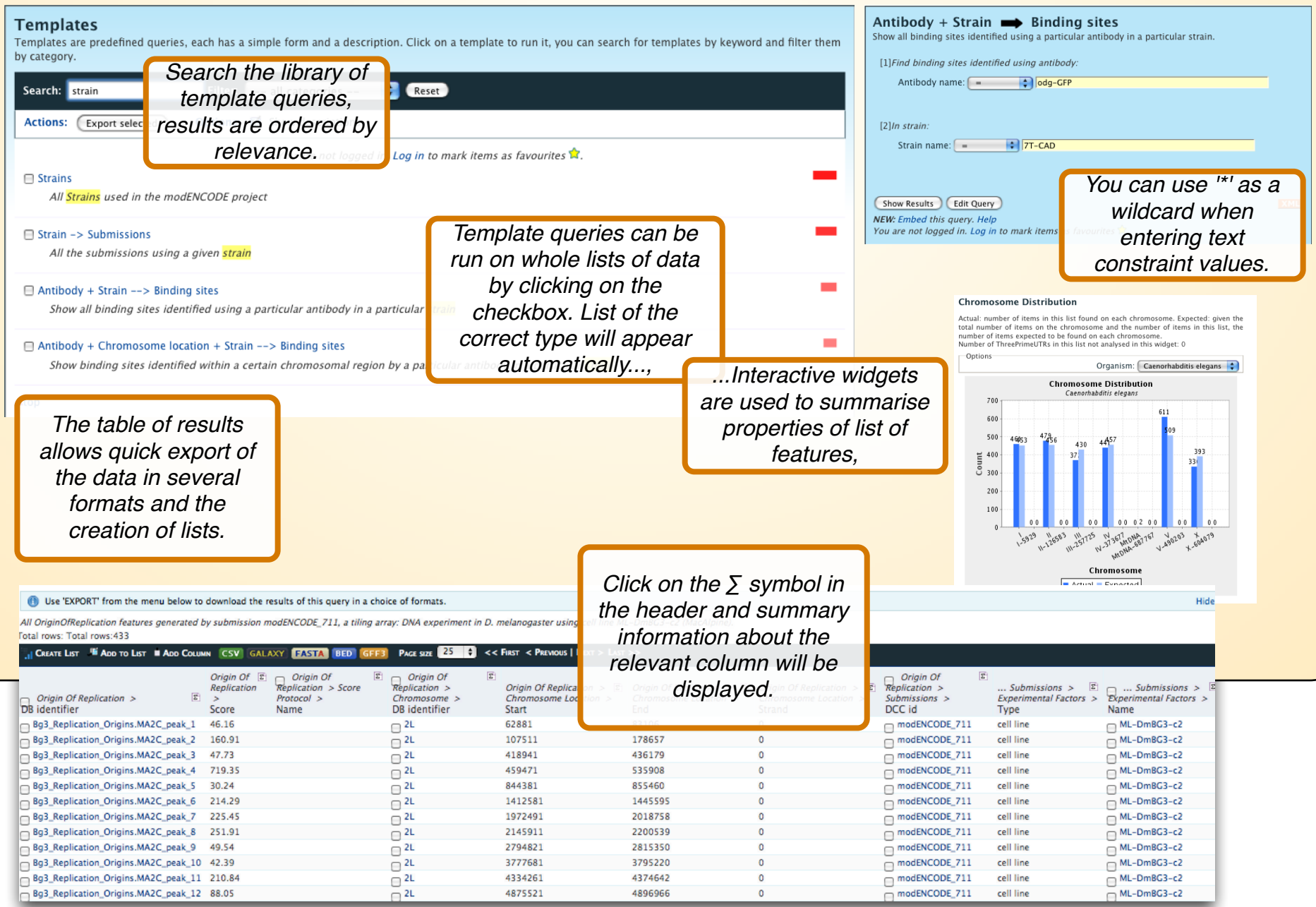
Web Service API

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**

UNIVERSITY OF CAMBRIDGE

OICR Ontario Institute for Cancer Research

BERKELEY LAB

genome.gov
National Human Genome Research Institute
National Institutes of Health