FlyMine 4.0
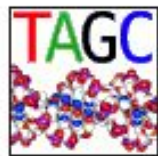

Genomics
Genome annotation

DrosDel
P-element
Insertions and
Deletions

RNAi
RNA interference

Proteins
Protein and proteomics data
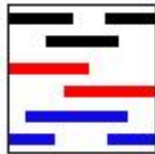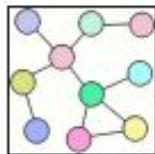
Gene Ontology

Gene Expression
ArrayExpress

Tiling Path
Microarray Tiling Primers

Disease
Human disease matches from
Homophila

Protein Interactions
IntAct

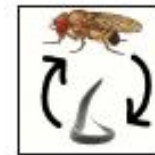**Binding Sites**
Protein binding sites

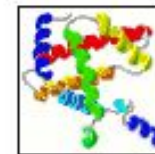**INDAC**
Long Oligos

Comparative Genomics
Orthologues
and Paralogues

Protein Structure
3-D protein
structures

An integrated database for
Drosophila and *Anopheles*

www.flymine.org

Outline

- What is FlyMine?

- Release 4.0 data sources

- Demonstration

www.flymine.org

UNIVERSITY OF CAMBRIDGE

What is FlyMine?

- A generic database allowing complex and arbitrary queries across diverse data sets

- Main focus: *Drosophila, Anopheles* and other insects (+ yeast, *C. elegans*)

- A genomics/proteomics focus

- Operates on lists

- Integrated viewing and analysis tools

UNIVERSITY OF
CAMBRIDGE

Motivation

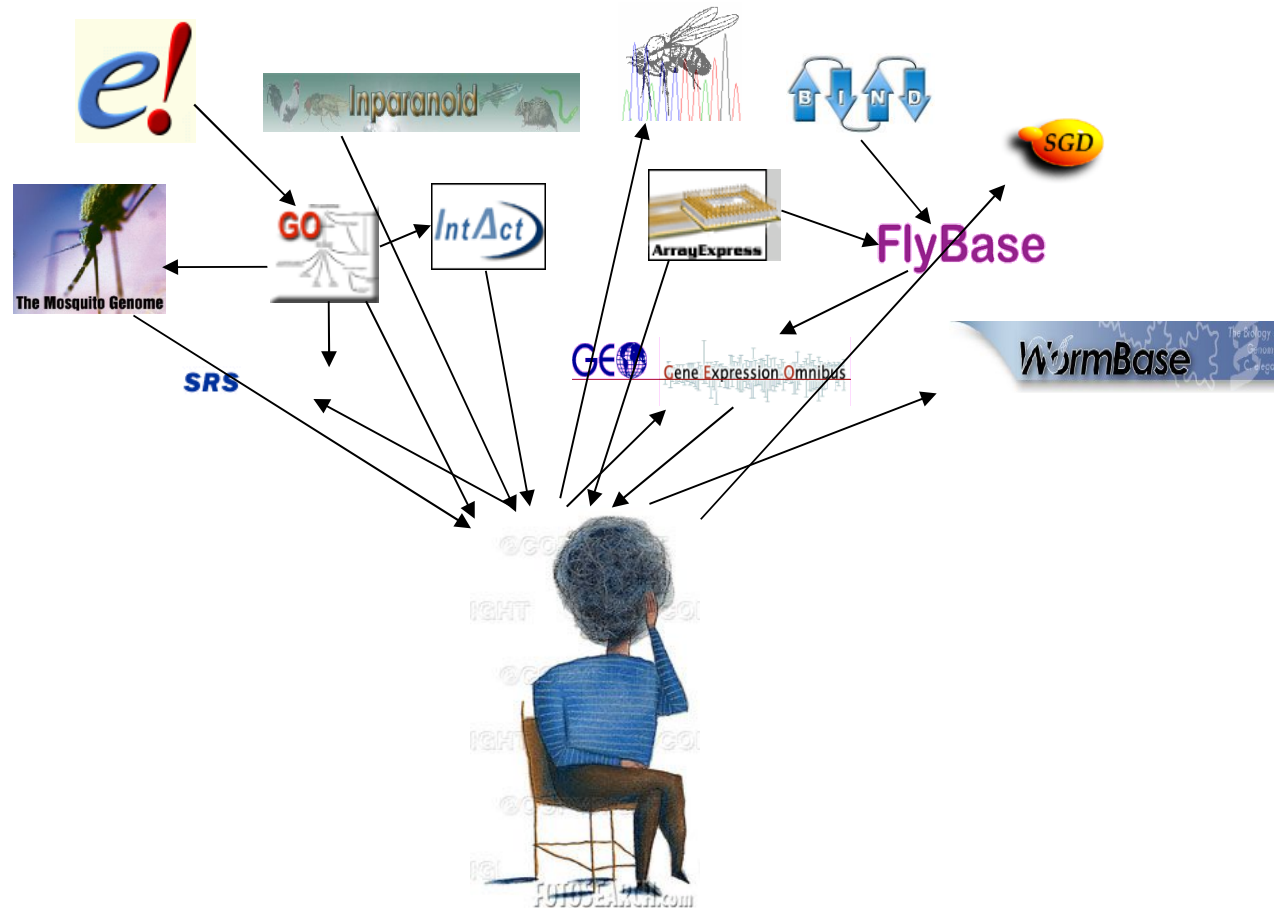- Massive amounts of experimental data (microarray, proteomics) being produced.

- Data quality can vary significantly : high-throughput datasets tend to be noisy.

- Stronger inferences can be made when data are combined from different sources.

UNIVERSITY OF CAMBRIDGE
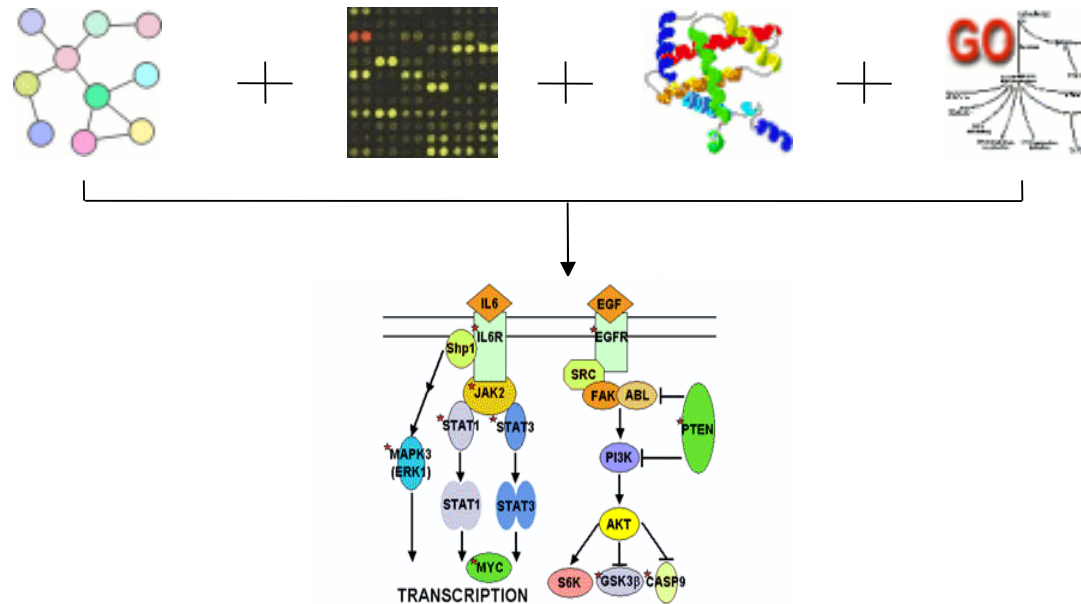
The Challenge of Integration

- Lots of databases
- Lots of different formats
- Need bioinformatics support

UNIVERSITY OF CAMBRIDGE

Creating an Integrated database

Genome Annotation

Orthologues

Expression Data

Disease Data

RNAi Phenotypes

Gene Ontology

Sequence Ontology

FlyMine Integrated Database

Protein Interactions

3D protein structures

Domain/Family data

Complex Query Builder

Browsing

Template Queries
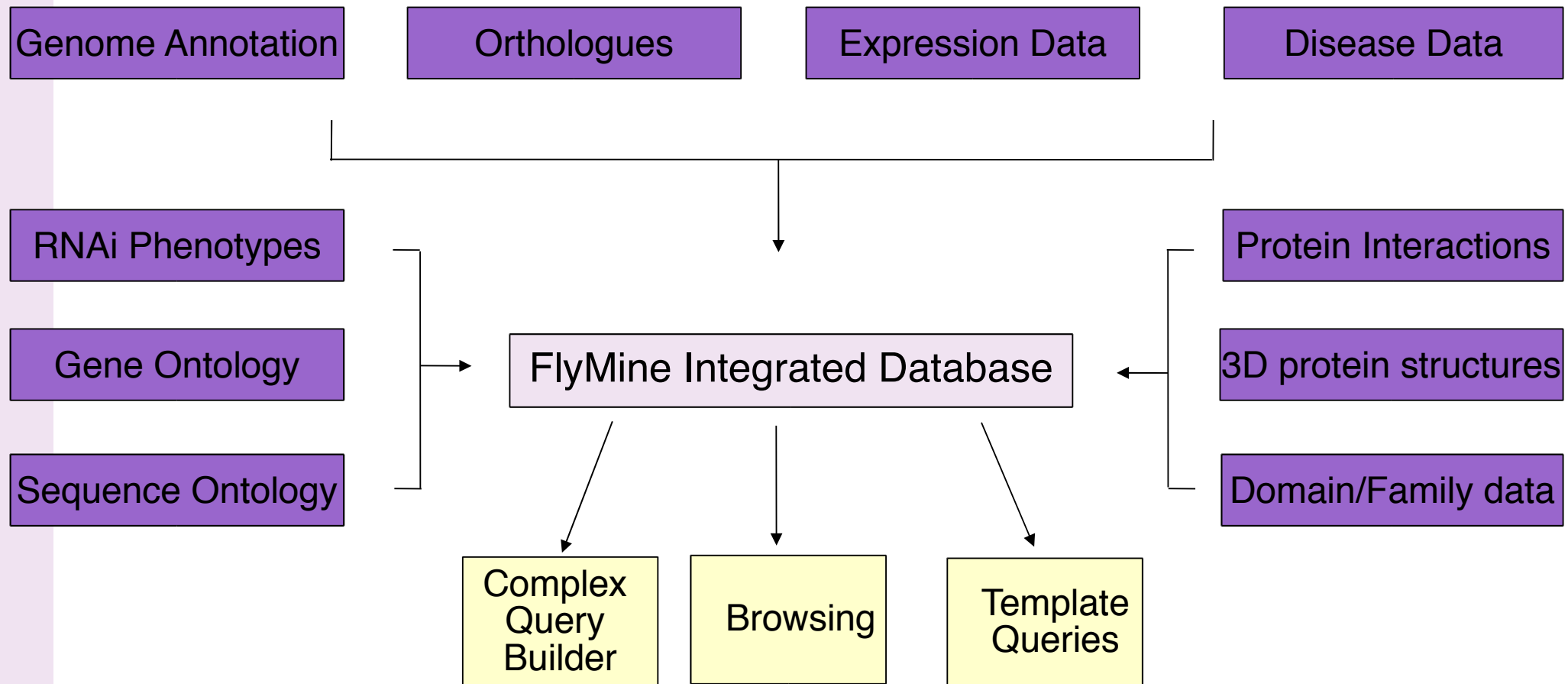
Easy to configure, easy to add more data

UNIVERSITY OF CAMBRIDGE

Release 4.0 Data Sources

- *D. melanogaster* genome annotation - **FlyBase**

- *A. gambiae* genome annotation - **Ensembl**

- **InParanoid** orthologues between multiple organisms

- **UniProt** protein data for multiple organisms

- Protein-protein interactions (multiple experiments) for *D. melanogaster* and *C. elegans* - **IntAct**

- Protein Family and Domain data for *D. melanogaster*, *A. gambiae* and *C. elegans* - **InterPro**

- *D. melanogaster* protein domain 3-D structure predictions - Kenji Mizuguchi, University of Cambridge

- RNAi phenotypes for *C. elegans* genes - **WormBase**

- **GO** annotations for *D. melanogaster*, *A. gambiae* and *C. elegans*

- DNase I footprint data for *D. melanogaster* - **FlyReg**

- Microarray gene expression data, including Arbeitman et al (2002) *D. melanogaster* time course experiment

- *D. melanogaster* **DrosDel** P-element insertion and deletion locations

- **Homophila** human disease to *Drosophila* gene data set

- INDAC microarray oligo set for *D. melanogaster* - **INDAC Consortium**

- *D. melanogaster* whole genome tiling amplimer set
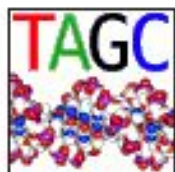
UNIVERSITY OF
CAMBRIDGE

Aspects

Genomics
Genome annotation

DrosDel
P–element Insertions and Deletions

RNAi
RNA interference

Proteins
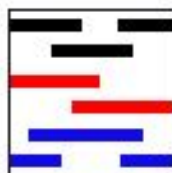Protein and proteomics data

Gene Ontology
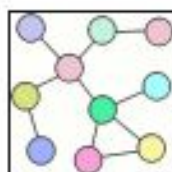
Gene Expression
ArrayExpress

Tiling Path
Microarray Tiling Primers

Disease
Human disease matches from Homophila
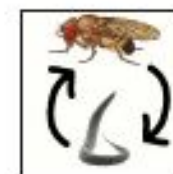
Protein Interactions
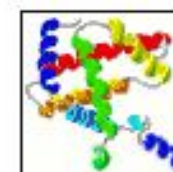IntAct

Binding Sites
Protein binding sites

INDAC
Long Oligos

Comparative Genomics
Orthologues and Paralogues

Protein Structure
3–D protein structures

www.flymine.org

UNIVERSITY OF CAMBRIDGE

Aspects

- Links to source databases and publications

- Access to bulk datasets

- Query start points

- Template queries

  - for simpler access to the data

  - *For a set of genes, search for the C. elegans orthologues and their RNAi phenotypes.*

  - adapt to make new queries

  - easy to add more – suggestions welcome

UNIVERSITY OF CAMBRIDGE

Releases

- Release cycle (content, functionality)

- Release 5.0 (Easter 2006)

  - *D. pseudoobscura* – **FlyBase, UniProt**

  - *A. mellifera* – **Ensembl, UniProt,**

  - **InParanoid** orthologues and **GO annotation** for more organisms

- Future development - *input welcome*

UNIVERSITY OF
CAMBRIDGE

Help and Feedback

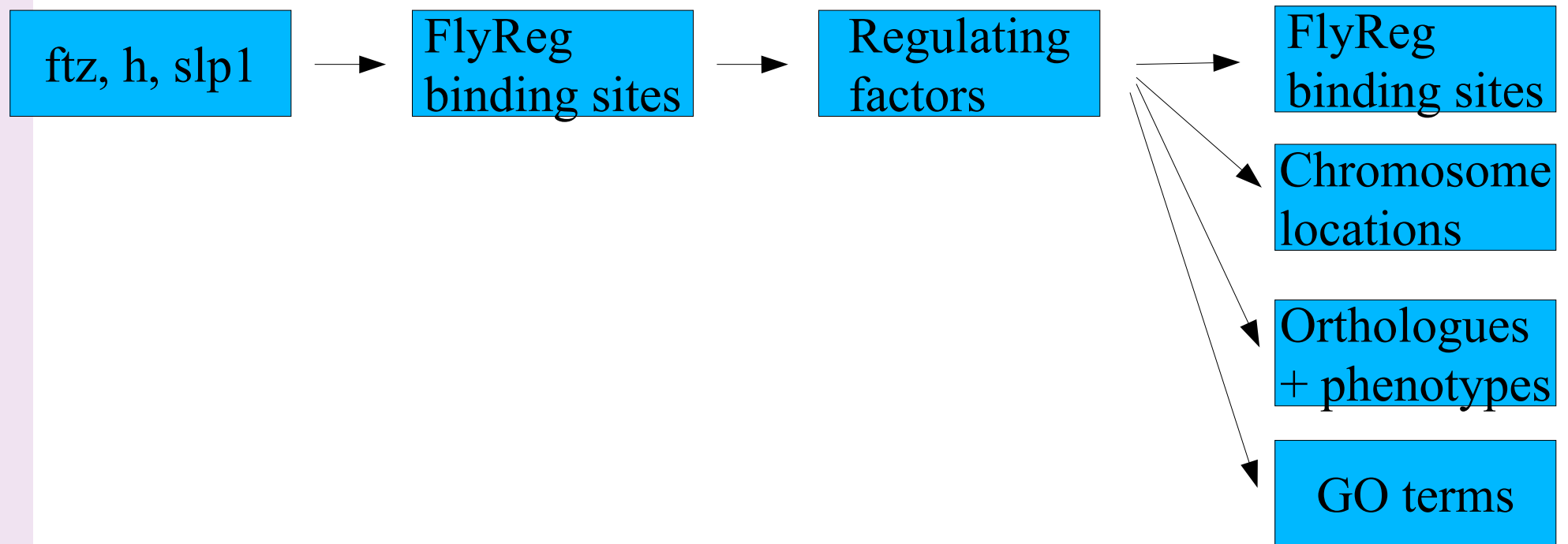Quick-start guide + Tutorials

Mailing lists (sign up sheets)

Support/ suggestions/ feedback welcome:

- Email :   support@flymine.org

- Feedback links on every FlyMine page

- Call us: +44 (0)1223 333 377

UNIVERSITY OF
CAMBRIDGE

Demonstation Example

ftz, h, slp1	→	FlyReg binding sites	→	Regulating factors

→ FlyReg binding sites

→ Chromosome locations

→ Orthologues + phenotypes

→ GO terms

Paste in list → Template query → Create list → More templates...

Acknowledgements

Richard Smith

Kim Rutherford

Matthew Wakeling

Tom Riley

Peter McLaren

Wenyan Ji

Rachel Lyne

Hilde Janssens

Florian Reisinger

François Guillier

Debashis Rana

Gos Micklem

More information at www.flymine.org

www.flymine.org

UNIVERSITY OF CAMBRIDGE