# SynBioMine by InterMine

# Integrated Data & Tools for Mining Synthetic Biology

Mike Lyne, Lalitha Sundaram, Fengyuan Hu, Jim Ajioka, Gos Micklem

Systems Biology Centre, University of Cambridge,
Tennis Court Road, Cambridge, CB2 1QR, UK

synbiomine@intermine.org
+44 1223 760262

## InterMine is:

- An established, integrated data warehouse system
  - adopted by many of the leading model organism databases:

Budding yeast (SGD), nematode (WormBase), zebrafish (ZFIN), mouse (MGI) and rat (RGD):

YeastMine: yeastmine.yeastgenome.org
WormMine: intermine.wormbase.org
ZfinMine: zmine.zfin.org/zebrafishmine
MouseMine: beta.mousemine.org
RatMine: ratmine.mcw.edu

The InterMine group also maintains FlyMine, an InterMine database for *Drosophila*.

## Why InterMine?

- Intuitive web interface
- Optimised Query Engine for searching across data
- Supports upload and analysis of lists of data
- Full programmatic access through REST-ful web services with client libraries in a range of languages
- Library of embeddable, graphical analysis and visualisation widgets

## SynBioMine's Data Sources

NCBI Gene, UniProt, PubMed, GO the Gene Ontology, IntAct, BioGRID

Likely future data types: pathways, expression, regulation, operons, enzyme activities, mutants and phenotypes

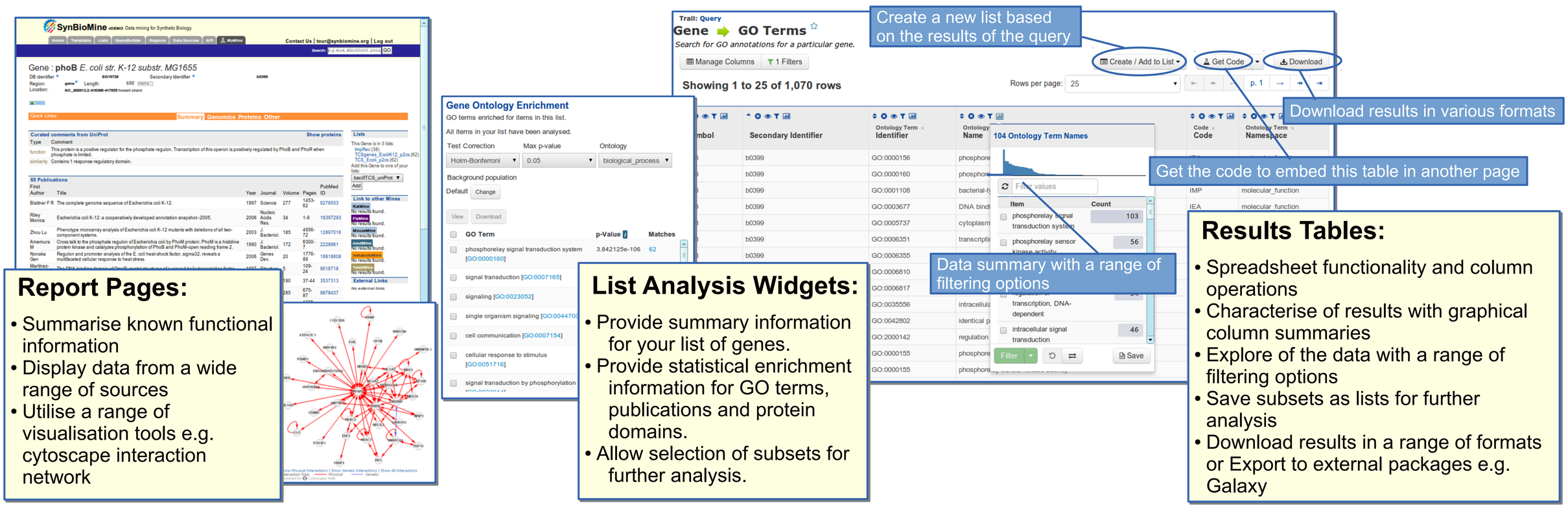
## Organisms

*E. coli* K-12 substr. MG1655
*E. coli* K-12 substr. W3110
*B. subtilis* subsp. subtilis str. 168

Further organisms and strains will be added as the project progresses

## SynBioMine's Web Interface

An intuitive web interface offers a range of features (list upload, predefined search forms, enrichment analysis and dynamic results tables) allowing researchers to build complex searches across domains of knowledge. Covering many different data sources and types, SynBioMine accepts a range of Identifiers including Gene symbol, Gene and Protein IDs and also synonyms.




### Report Pages:

- Summarise known functional information
- Display data from a wide range of sources
- Utilise a range of visualisation tools e.g. cytoscape interaction network

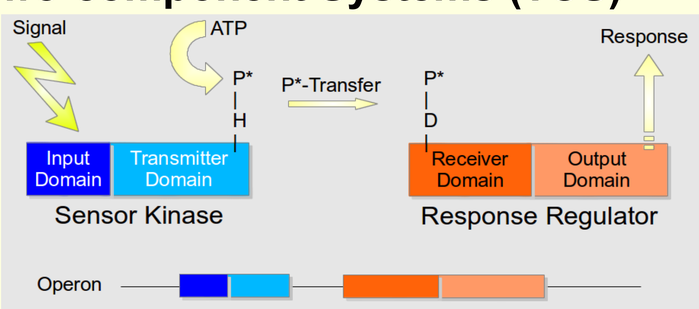
### List Analysis Widgets:

- Provide summary information for your list of genes.
- Provide statistical enrichment information for GO terms, publications and protein domains.
- Allow selection of subsets for further analysis.

### Results Tables:

- Spreadsheet functionality and column operations
- Characterise of results with graphical column summaries
- Explore of the data with a range of filtering options
- Save subsets as lists for further analysis
- Download results in a range of formats or Export to external packages e.g. Galaxy

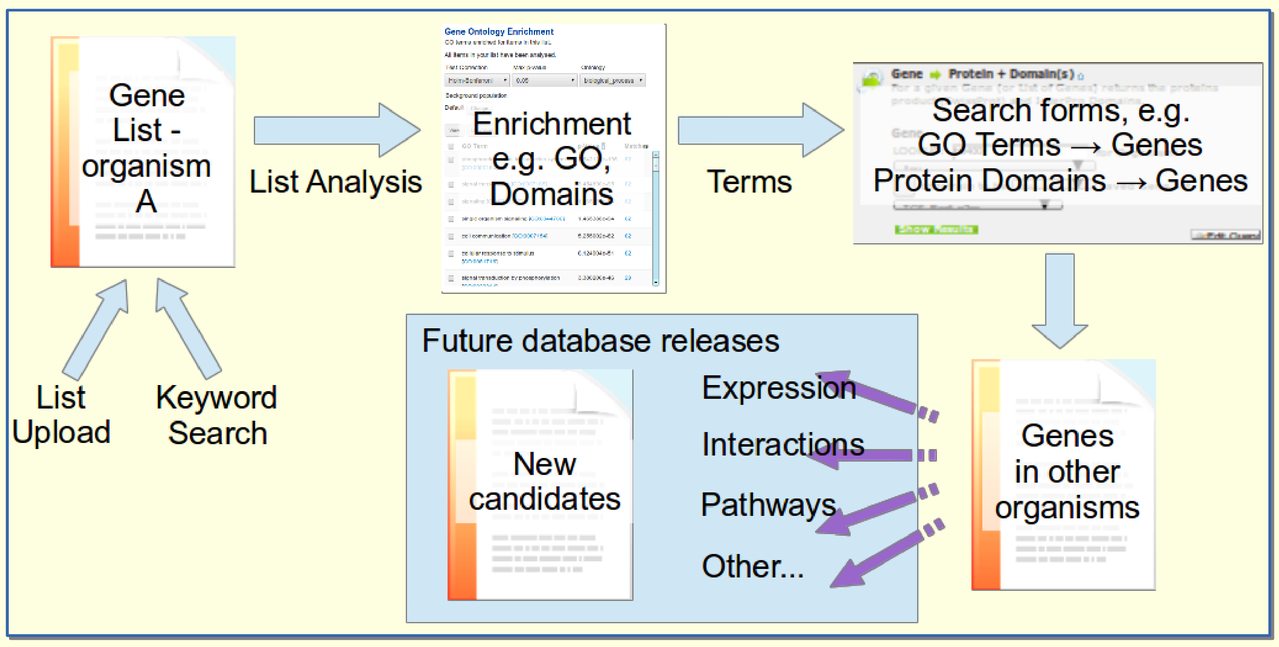
## Two-component Systems (TCS)




Two-component systems are common signal transduction pathways which, at a basic level, involve two multi-domain proteins. The first, a histidine protein kinase (HPK), is activated upon receiving an environmental stimulus. The second, a response regulator (RR), receives a phosphoryl group, transferred from the phosphorylated HPK, then mediates phosphorylation-dependent effects within the cell.

Chimeric TCS are likely to provide many of the component parts for the assembly of synthetic genetic networks

## Use Case – Mining for TCS with SynBioMine




Starting with a list of genes in one organism, automated list analysis identifies a set of enriched terms. These terms serve as input for predefined search forms to identify corresponding genes from other organisms.

Cross-reference against additional data types allows both refinement and expansion of candidate gene sets.

UNIVERSITY OF CAMBRIDGE
FLOWERS Consortium
EPSRC Pioneering research and skills
InterMine