InterMineR

R Client Library for InterMine

Rachel Lyne
Department of Genetics
University of Cambridge

UNIVERSITY OF CAMBRIDGE

InterMine

What is InterMine?

Disparate data in, unified data out

WEBAPP

InterMine

API & CLIENTS

Ruby | Java | Perl

R | Javascript | Python
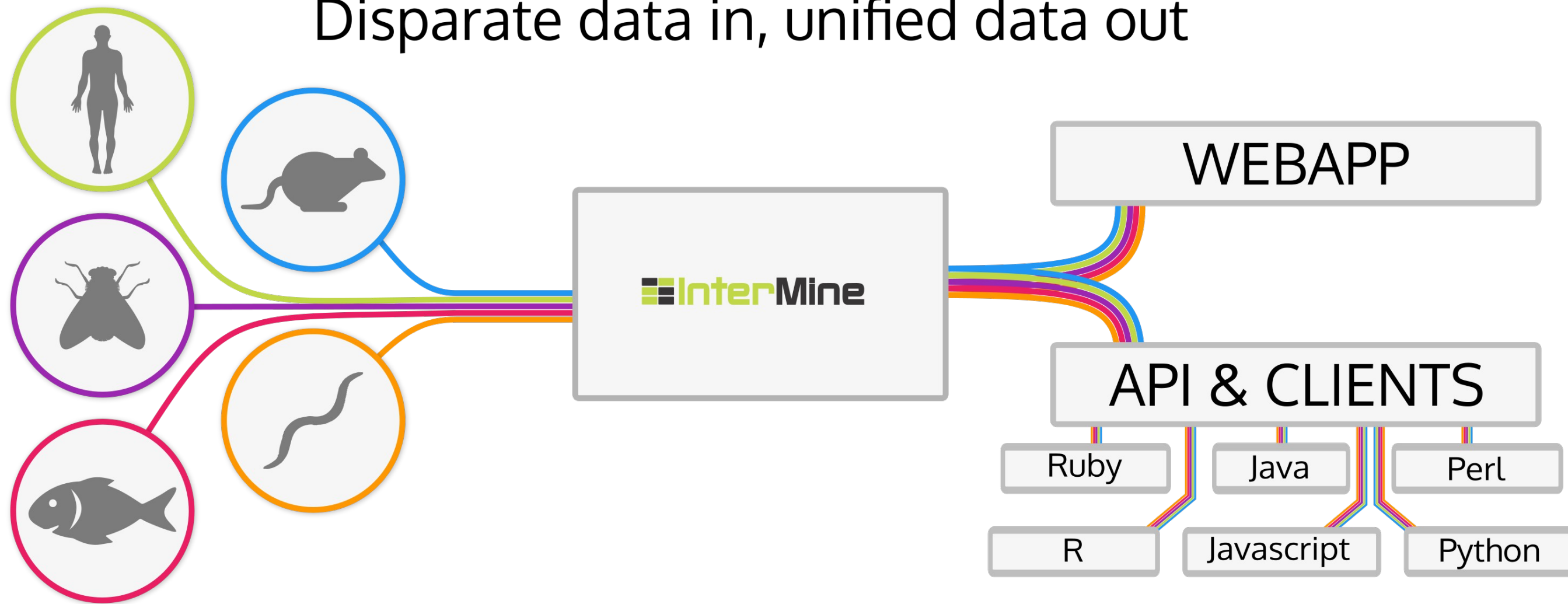
Model organism images Designed by Freepik and distributed by Flaticon

Who Uses InterMine?

SynbioMine
FlyMine
RatMine
Wheat3BMine
ZebrafshMine
HumanMine
PlanMine
MedicMine
WormMine
MouseMine
HymenopteraMine
modMine
TargetMine
YeastMine
BovineMine
PhytoMine
ThaleMine
ChickpeaMine
XenMine
PeanutMine
SoyMine
LegumeMine
PeanutMine
CowpeaMine

UNIVERSITY OF CAMBRIDGE

InterMine

Data

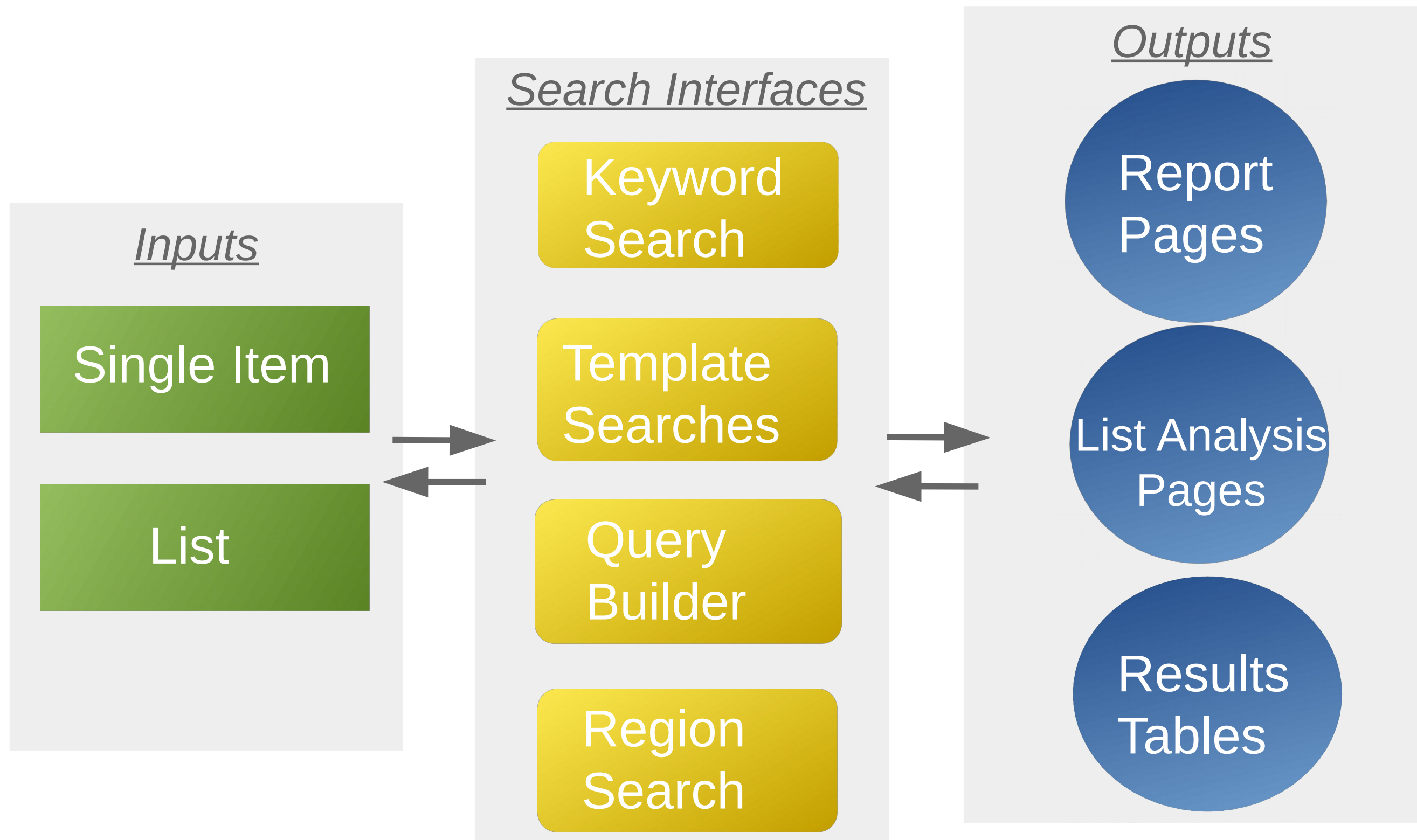TRANSCRIPTS ORTHOLOGUES GENE ONTOLOGY GENES UTRs INTERACTIONS EXONS PATHWAYS PROTEINS Protein Domains GWAS REGULATORY VARIANTS SNPs MICROARRAY RNA-seq ALLELES PHENOTYPES EXPRESSION DISEASE

UNIVERSITY OF CAMBRIDGE

InterMine

The Web Interface

Inputs

- Single Item
- List

Search Interfaces

- Keyword Search
- Template Searches
- Query Builder
- Region Search

Outputs

- Report Pages
- List Analysis Pages
- Results Tables

Google Summer of Code 2017

June – August - Konstantinos Kyritsis

Proposal

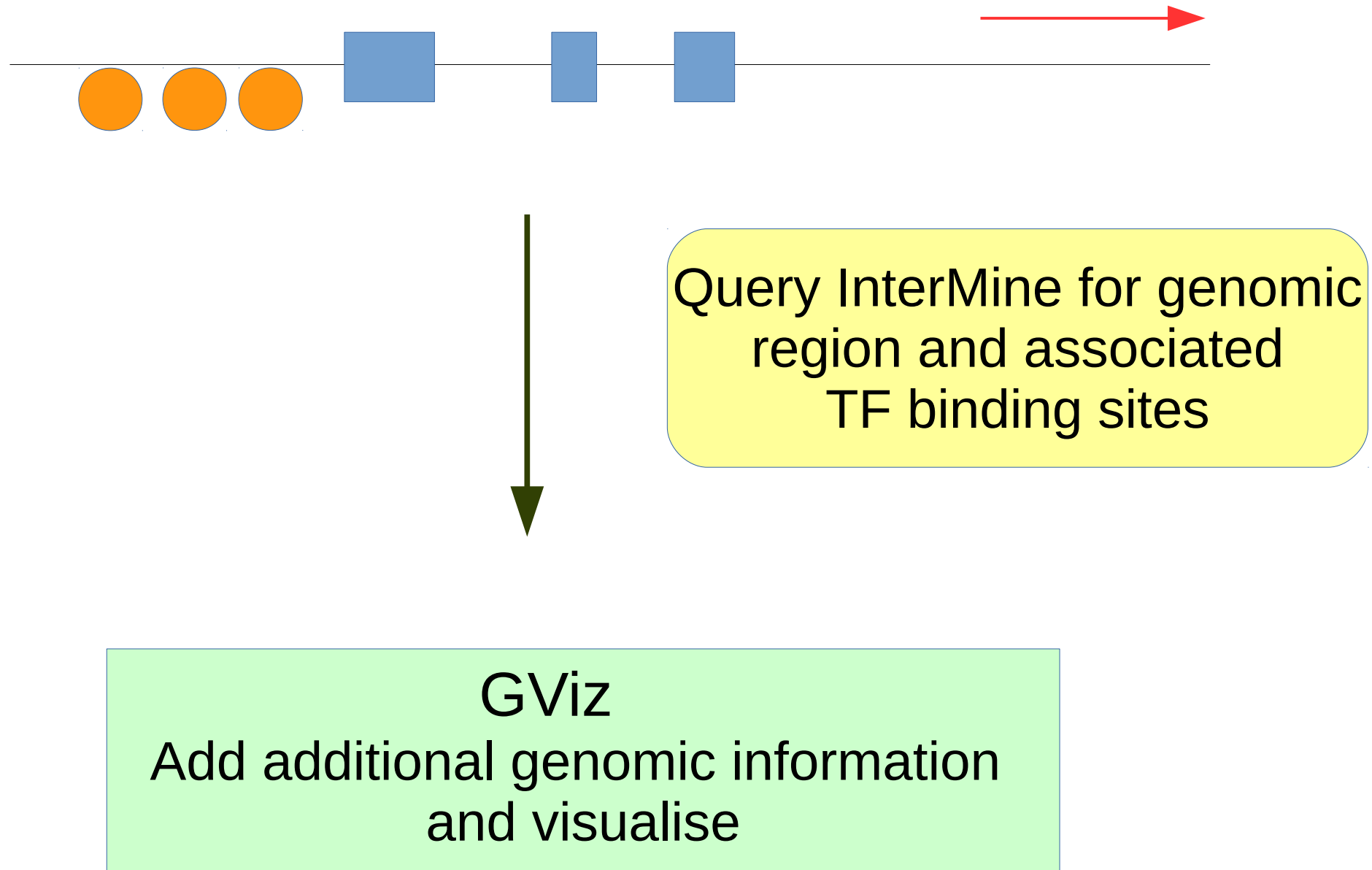Add to Bioconductor

Add Visualisation

Add Enrichment + visualisation

Add Query Builder

What can InterMineR do?

- Data model

- Data sets

    - Which data is in this InterMine?

- Templates

    - List of available templates
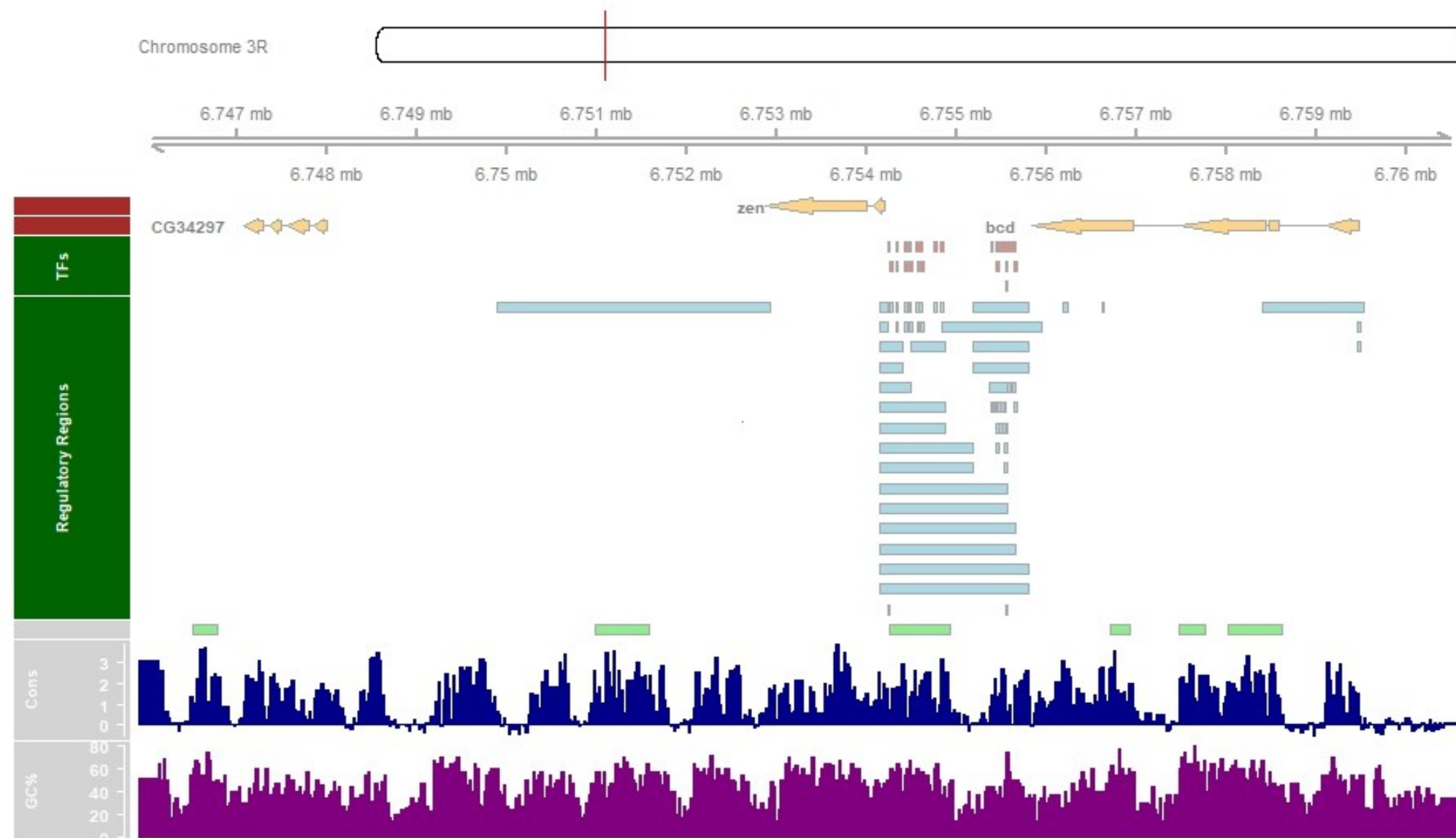
- Execute, modify templates

- Queries

UNIVERSITY OF
CAMBRIDGE

InterMine

Explore a genomic region....

Query InterMine for genomic
region and associated
TF binding sites

GViz
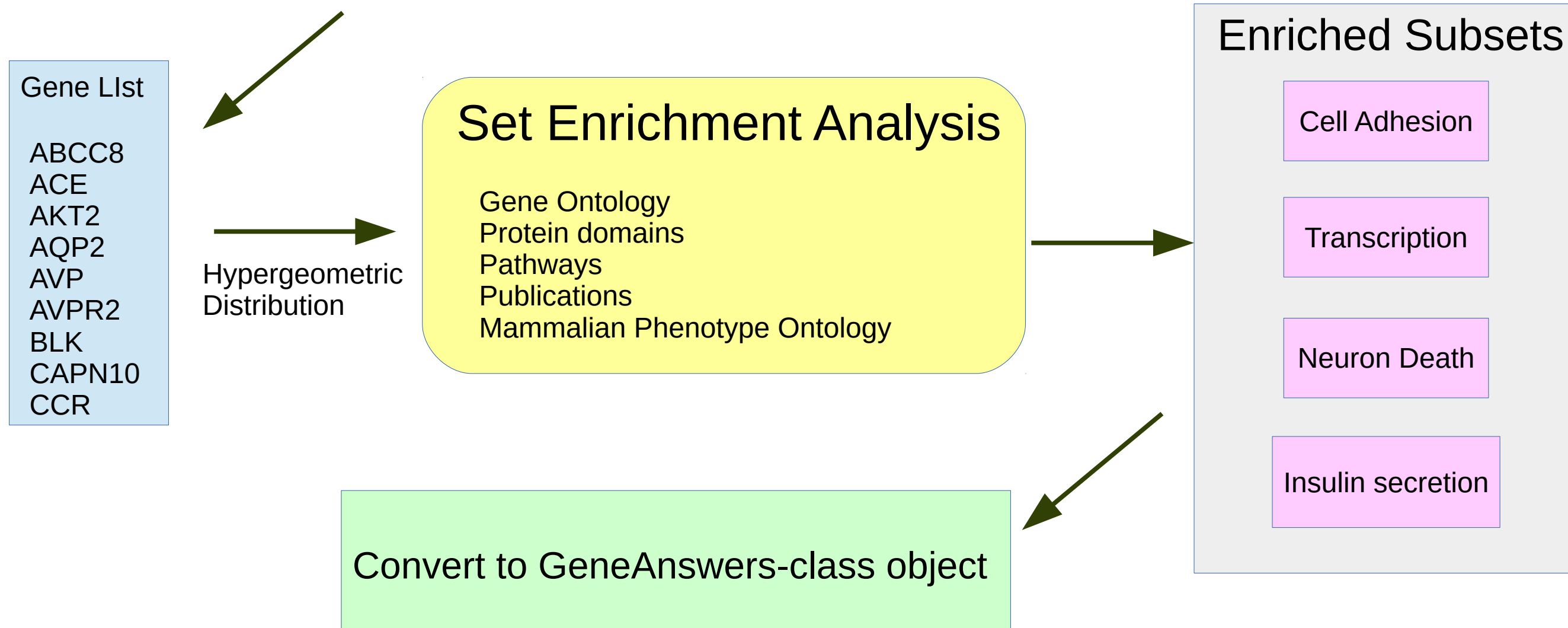Add additional genomic information
and visualise

GeneAnswers is an R package that provides statistical and network visualization functions in order to explore possible relationships between a group of genes and a list of categories (e.g. Gene Ontology) (Huang et al. 2014).

UNIVERSITY OF
CAMBRIDGE

InterMine

Explore a genomic region....

The genomic region around the *Drosophila* zen gene

Analyse a set of genes....

InterMine

Gene LIst

ABCC8
ACE
AKT2
AQP2
AVP
AVPR2
BLK
CAPN10
CCR

Hypergeometric
Distribution

Set Enrichment Analysis

Gene Ontology
Protein domains
Pathways
Publications
Mammalian Phenotype Ontology

Enriched Subsets

Cell Adhesion

Transcription

Neuron Death

Insulin secretion
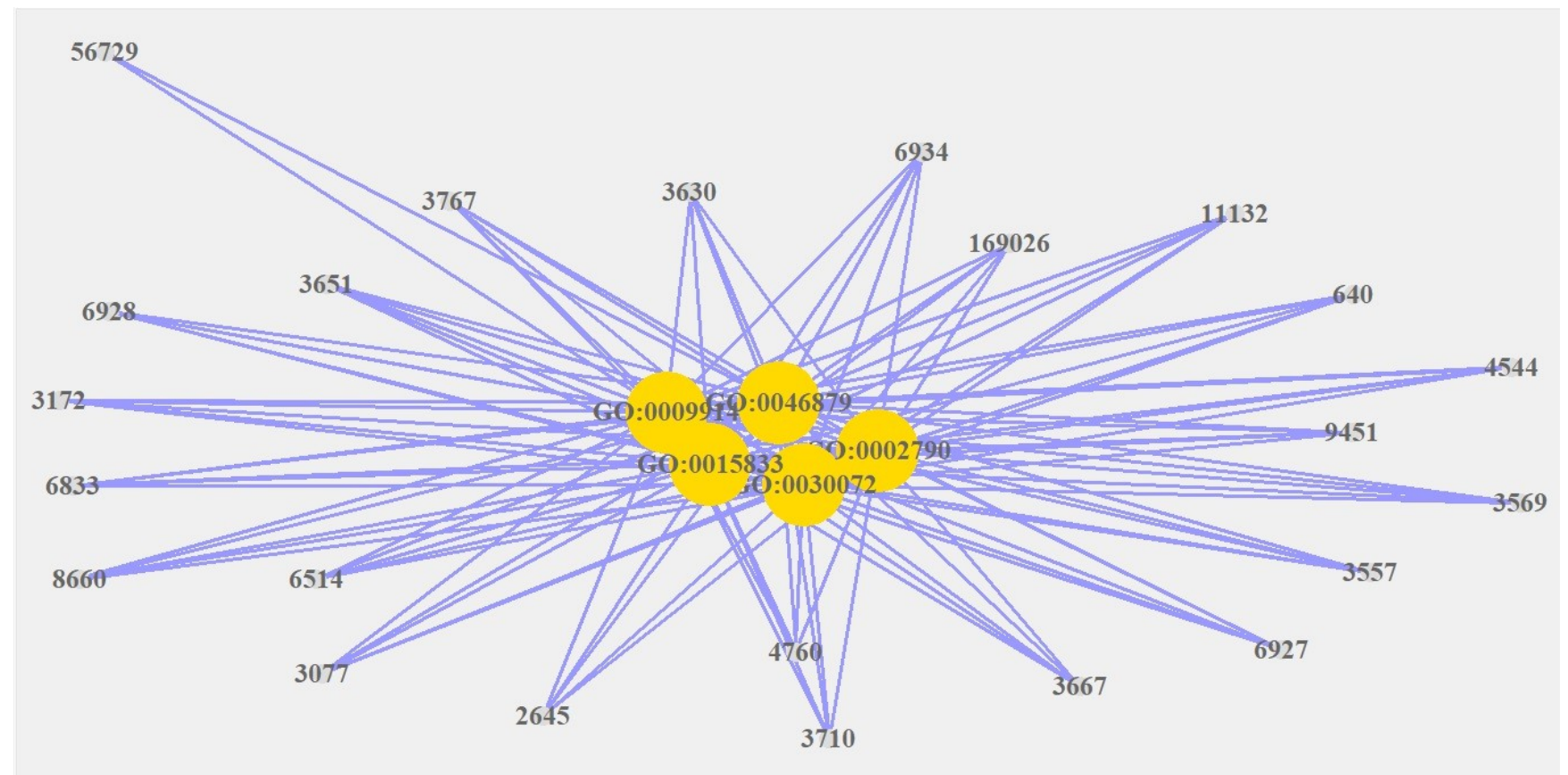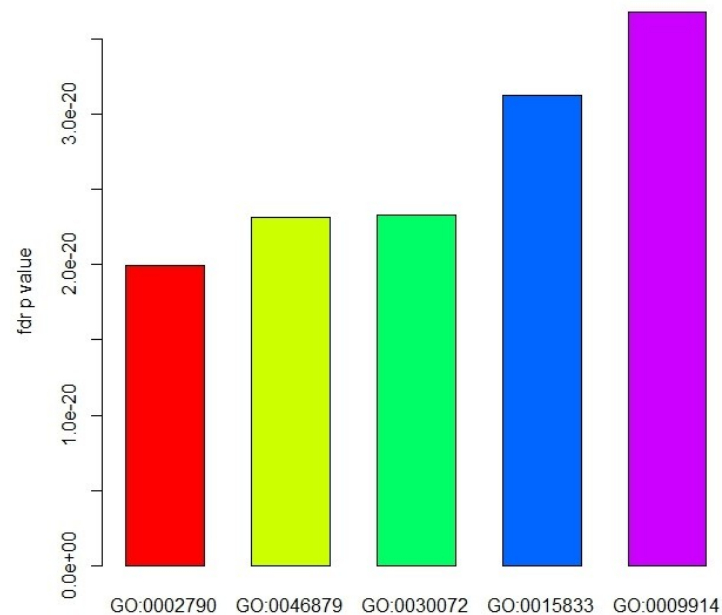
Convert to GeneAnswers-class object

GeneAnswers is an R package that provides statistical and network visualization functions in order to explore possible relationships between a group of genes and a list of categories (e.g. Gene Ontology) (Huang et al. 2014).

UNIVERSITY OF
CAMBRIDGE

InterMine

Enrichment Visualisation
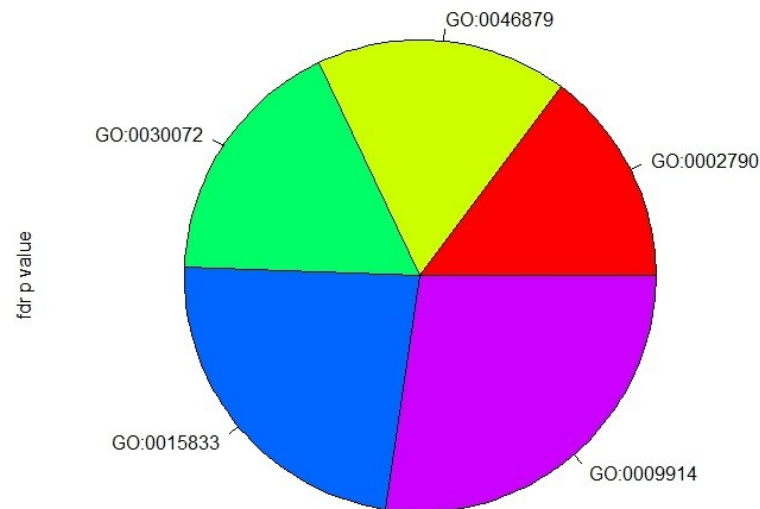
Fascilitating integration with BioConductor packages......

- convertToGRanges()

  - Scalable data structures for annotated genomic ranges

- convertToRangedSummarizedExperiment()

  - Facilitate the analysis of gene expression data
    and associated annotations

Future work

Shiny app to build queries with UI

Generate R queries in InterMine UI

More features

    - Any suggestions!

UNIVERSITY OF CAMBRIDGE

Credits

Main Developer - Konstantinos Kyritsis

Mentors - Julie Sullivan and Rachel Lyne

GSoC coordinator - Yo Yehudi

PI - Gos Micklem

Analyse a set of genes....

Enrichment

Given a list, give me enrichment for a chosen annotation

Enrichment visualisation

InterMine's enrichment tool - hypergeometric distribution with choice of multiple test correction.

Convert InterMineR enrichment analysis results to a GeneAnswer-class object .

GeneAnswers is an R package that provides statistical and network visualization functions in order to explore possible relationships between a group of genes and a list of categories (e.g. Gene Ontology) (Huang et al. 2014).

UNIVERSITY OF CAMBRIDGE

InterMine

Explore a genomic region.....

- Query InterMine for genomic locations

- Query InterMine for TF binding sites and regulatory regions for these regions

- Use GeneViz to pull back other interesting data - e.g GC content, CpG islands and conservation score

- Use GViz to visualize the region and features.

UNIVERSITY OF CAMBRIDGE

InterMine

The InterMine Team

Gos Micklem (PI)

Julie Sullivan

Sergio Contrino

Joshua Heimbach

Daniela Butano

Yo Yehudi

Justin Clark-Casey

Rachel Lyne (rachel@intermine.org)