

Developing an application ontology for annotation of experimental variables across high-throughput data sets

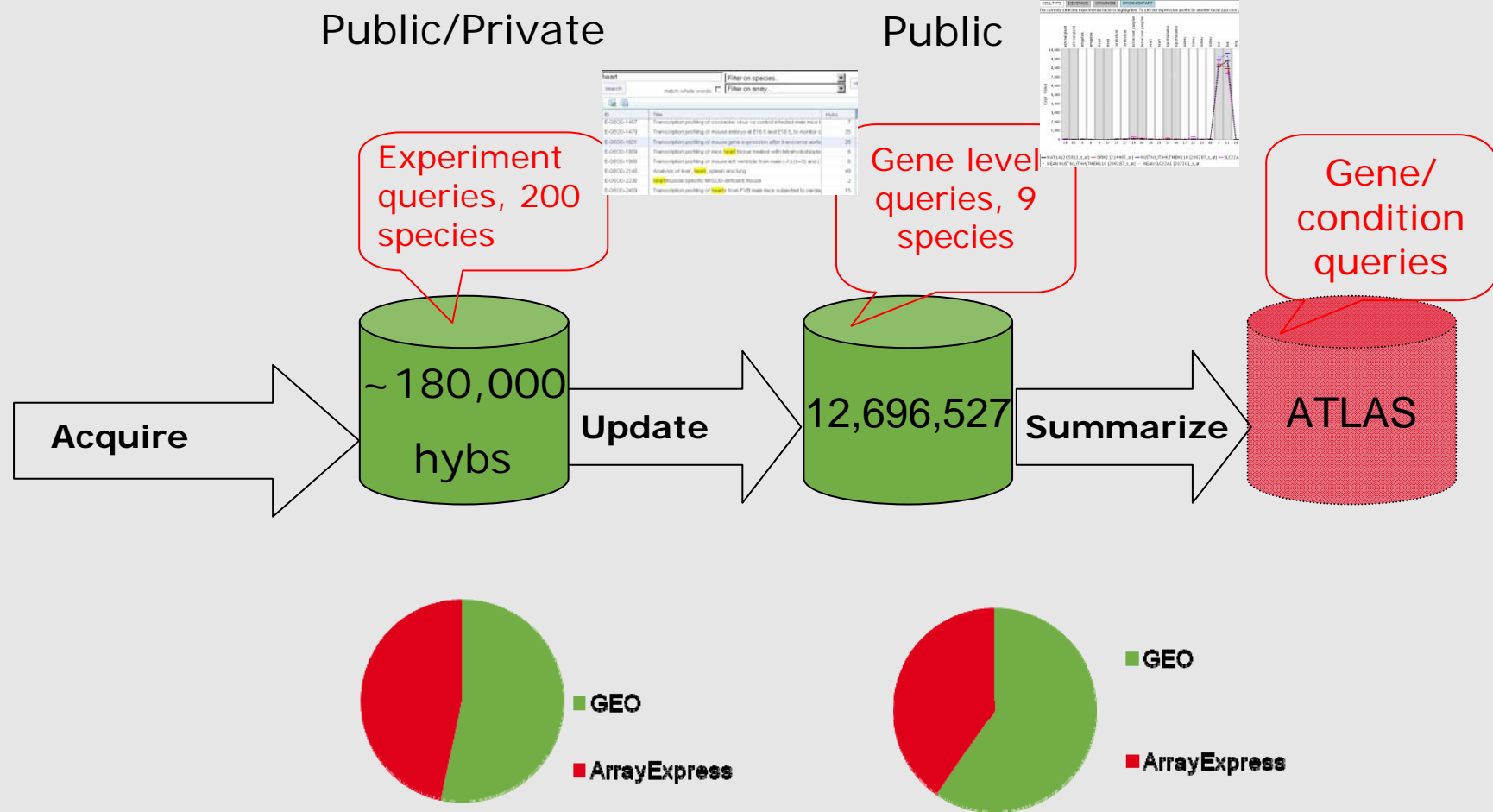
James Malone
malone@ebi.ac.uk



Overview

- Motivation
 - Our use cases
 - Annotating HTP experimental data
- Methodology for creating the ontology
 - Semi-automated mapping and manual curation
- Current ontology usage
- Future use and challenges

ArrayExpress State of the Art

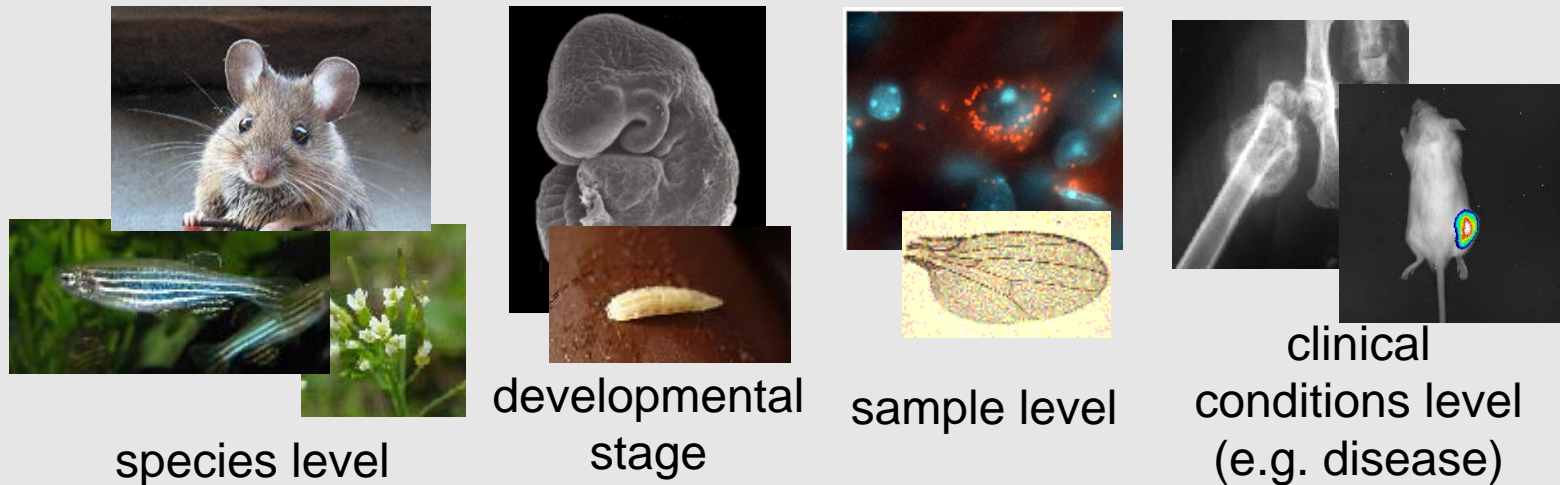


Our Use Cases

- Query support (e.g, query for 'cancer' and get also 'leukemia')
- Over-representation analysis in groups of samples (analogous to the use of GO terms in over-representation analysis in groups of genes)
- Data visualisation – e.g., presenting an ontology tree to the user of what is in the database
- Data integration by ontology terms – e.g., we assume that 'kidney' in independent studies roughly means the same, so we can count how many kidney samples we have in the database
- Intelligent template generation for different experiment types in submission or data presentation
- Summary level data

Scope of Experimental Factor Ontology (EFO)

- Modelling all of the experimental factors that are currently present in the ArrayExpress repository
- Experimental factors are variable aspects of an experiment design which can be used to describe an experiment
- Scope is primarily determined by data currently held in ArrayExpress



'Experimental Factors'

E-MEXP-114 Transcription profiling of hypothalamus, liver, kidney, ovaries and testis from male and female humans and mice

Gene Expression Profile for (Acsm2)

ORGANISM ORGANISMPART SEX

Gene Properties [expand]

Gene name

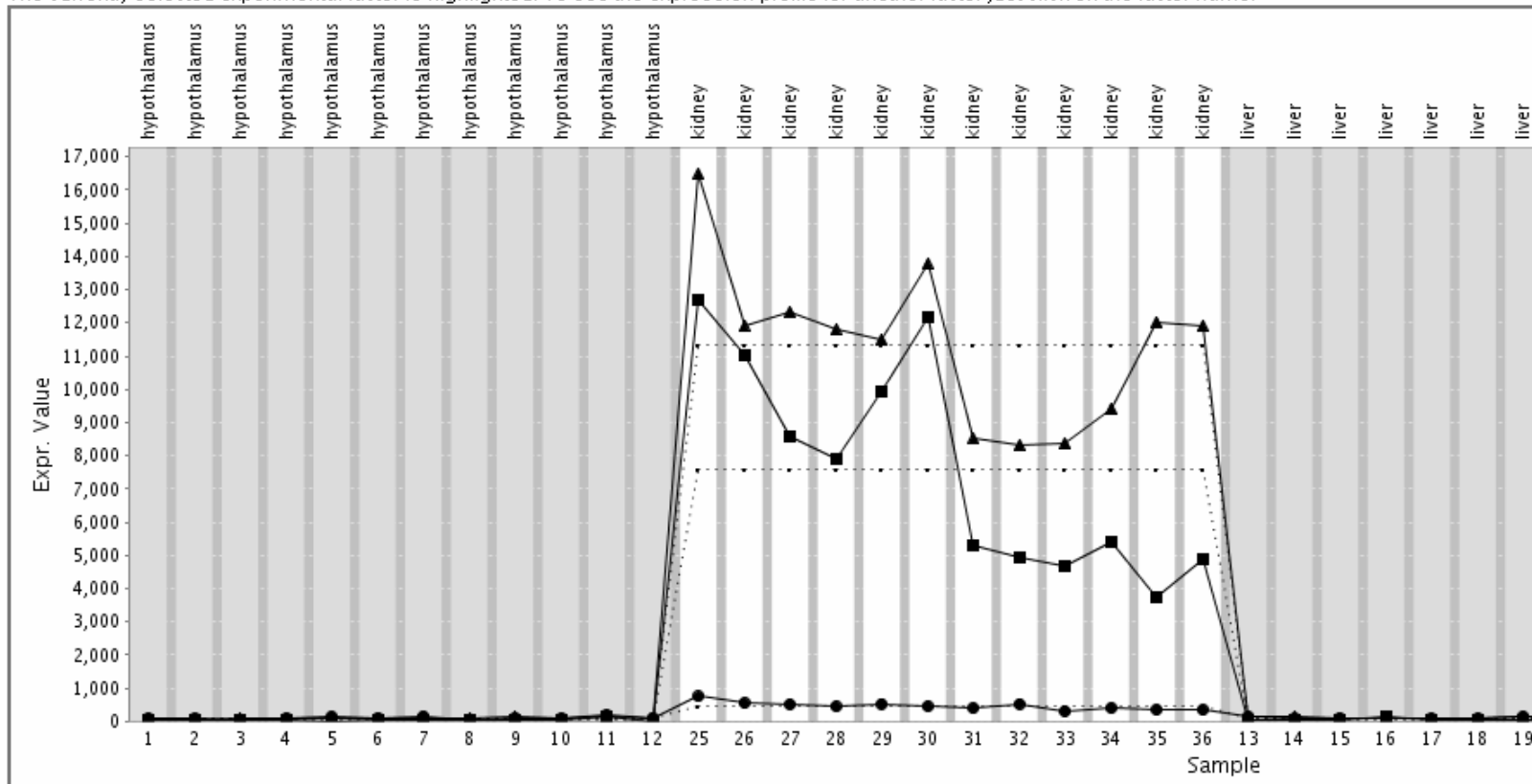
Acsm2

1427223_a_at
1427224_a_at
1456190_a_at
BC031140

similarity search

- Gene Ontology: butyrate-CoA ligase activity, catalytic activity, metabolic process

The currently selected experimental factor is highlighted. To see the expression profile for another factor just click on the factor name.



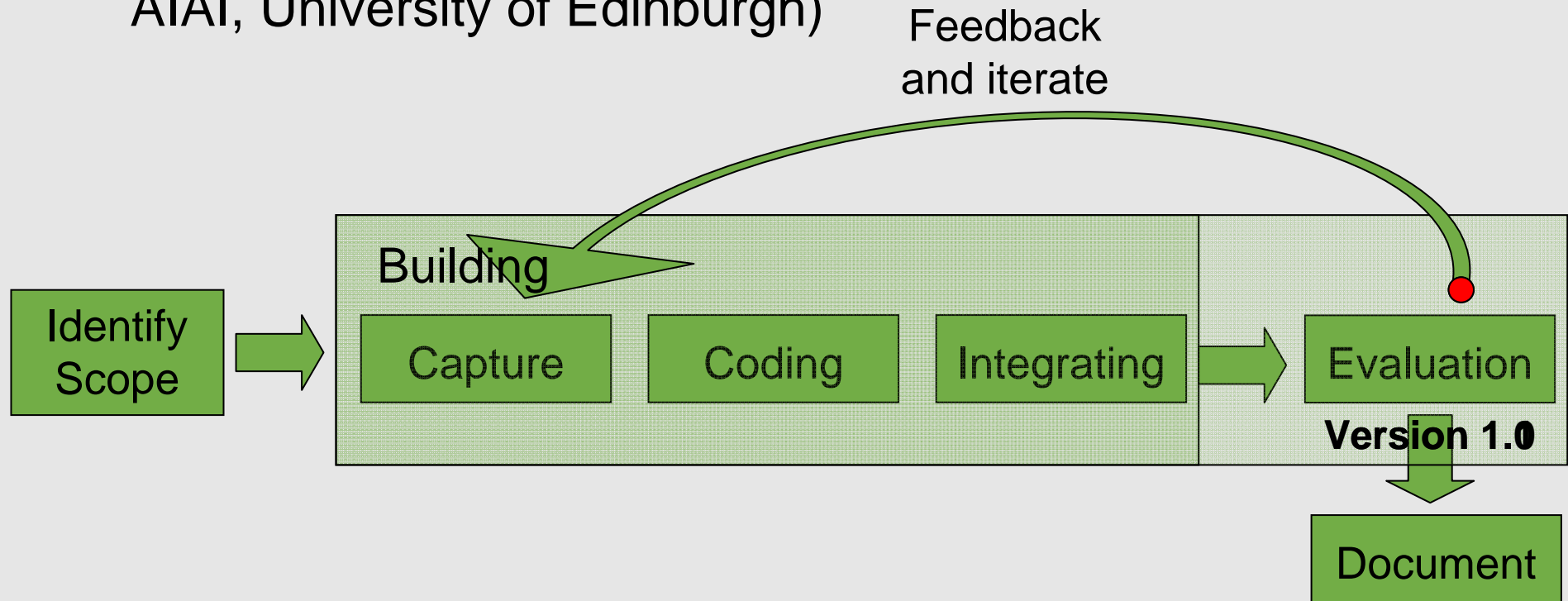
Done

Our Needs

- We need an ontological representation that:
 - Serves our bioinformatic use cases
 - Reuse, where appropriate, of existing resources
 - Integrates with such resources with audit trail
 - Does not involve user having to familiarise with many ontologies and interfaces
 - Do it all whilst considering trade off between quickest build vs most accurate and extensible implementation

Our Approach

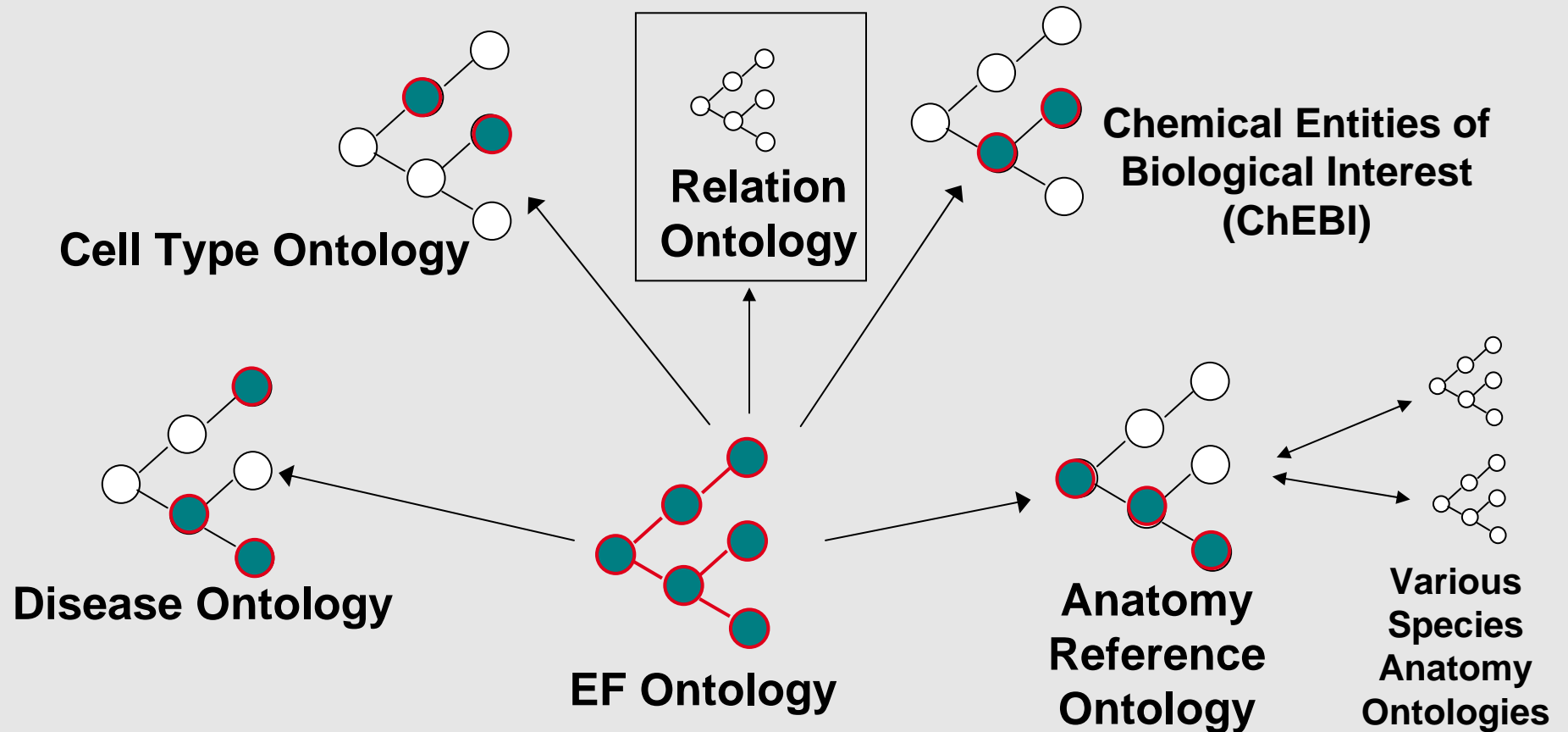
- Using Uschold and King's Methodology (developed at AIAI, University of Edinburgh)



And so on....

Semantic Roadmap

- Position of the ArrayExpress Experimental Factor Ontology in the 'bigger picture'
- Key is orthogonal coverage, reuse of existing resources and shared frameworks



Experimental Factor Ontology (EFO)

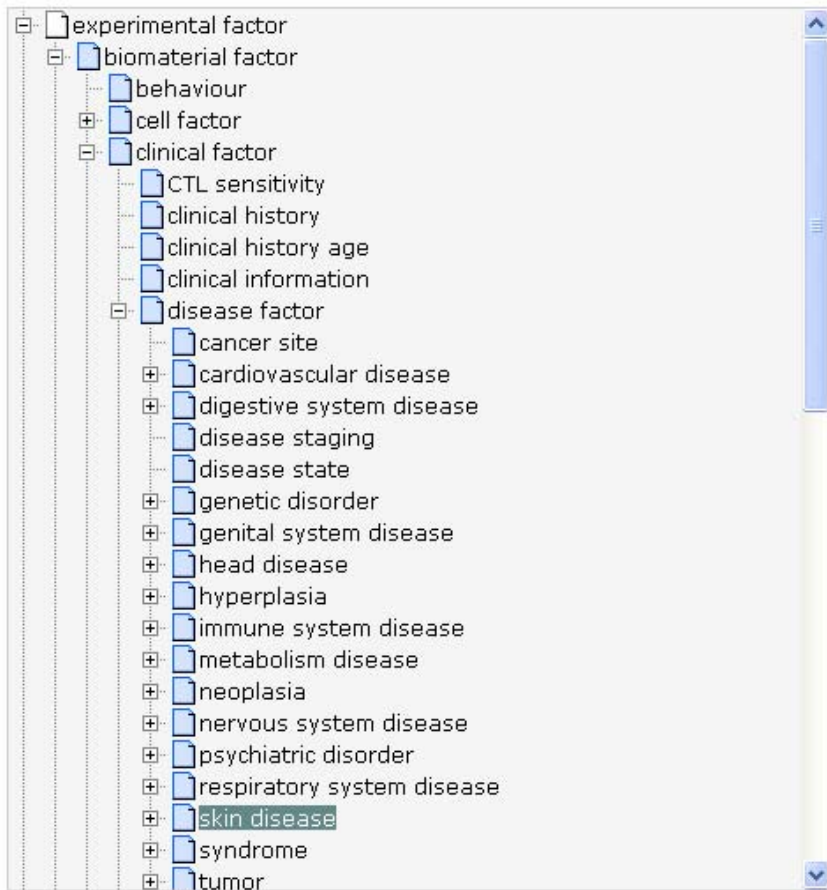


- Home
- Documentation
- Project
- Publications
- Developer Resources
- Download
- Implementation
- Overview
- Workflow
- Web Service
- Documentation
- Contact Us
- Acknowledgements

2007: Maintenance release
Documentation has been updated to include more dependencies. Note that unless otherwise indicated, newer versions of given services should work out of the box.
2007: Maintenance release
This is a new release of EFO available, which includes mostly maintenance fixes and code improvements.

OLS - Ontology Lookup Service

EFO Ontology Browser



Help (hide)
Double-click a term to see its children. The ontology browser is populated dynamically. If there are many children for a given term, there may be a small delay while the browser fetches. **Click** to highlight a term to see any information associated with it. **Hover** over a term to see its relation with its immediate parent. Root terms will not display any relational information.

Relations

Term Information
ID: http://www.ebi.ac.uk/experimentalfactors.owl/EFO_0000701 Zoom
Name: skin disease

Associated information

experimentalfactors:definition	Any deviation from the normal structure or function of the skin or subcutaneous tissue that is manifested by a characteristic set of symptoms and signs.
experimentalfactors:definition_citation	NCI Thesaurus: C3371
experimentalfactors:definition_editor	James Malone

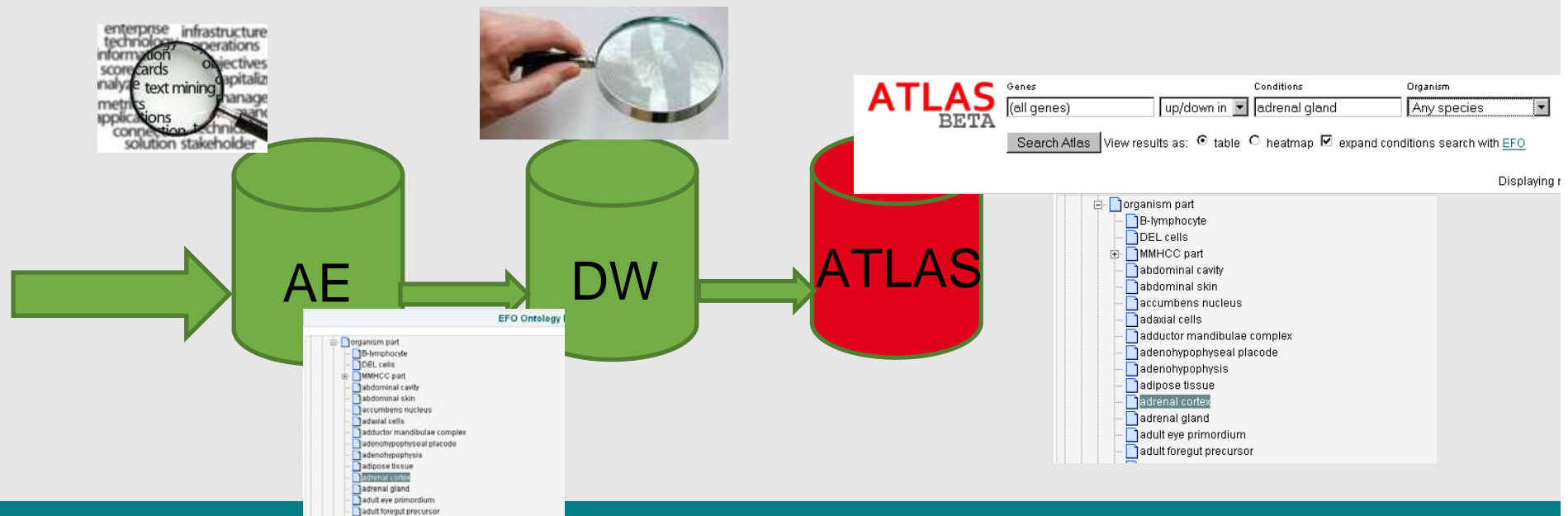
Term Hierarchy
Paths to Root: Child relationships:
http

Legend:

- is a
- develops from
- part of
- other

Application Ontology Status Quo

- Text mining at data acquisition
- Tuned for queries, structured for use in ArrayExpress GUI
- Multi-species aspect



EFO in Atlas (www.ebi.ac.uk/microarray-as/atlas)

home | [about the project](#) | [faq](#) | [feedback](#) | [blog](#) | [help](#)

ATLAS BETA

Genes: (all genes) Conditions: up/down in ▼ cancer Organism: Homo sapiens ▼

Search Atlas View results as: table heatmap expand conditions search with [EFO](#)

Your conditions query was expanded via [EFO](#) to: cancer sarcoma melanoma chordoma carcinoma

Displaying results 1-318 of ~318 found in the atlas

Experiment	Description	Factor Value (Factor)	Gene Name	Gene Id	Organism	P-value	AEW	...
E-GEOD-2712	Transcription profiling of human Clear cell sarcoma of the kidney (CCSK) vs Wilms tumors (WT) and fetal kidney samples to address diagnostic challenges of CCSK	clear cell <i>sarcoma</i> of the kidney (diseasestate)	WT1	ENSG00000184937	Homo sapiens	5.62e-28	More...	
E-GEOD-3189	Transcription profiling of human tissue samples to identify novel genes associated with malignant melanoma but not benign melanocytic lesions	malignant <i>melanoma</i> (diseasestate)	NFIB	ENSG00000147862	Homo sapiens	4.05e-26	More...	
E-GEOD-3189	Transcription profiling of human tissue samples to identify novel genes associated with malignant melanoma but not benign melanocytic lesions	malignant <i>melanoma</i> (diseasestate)	PHACTR1	ENSG00000112137	Homo sapiens	1.68e-25	More...	
E-GEOD-3189	Transcription profiling of human tissue samples to identify novel genes associated with malignant melanoma but not benign melanocytic lesions	malignant <i>melanoma</i> (diseasestate)	NFIB	ENSG00000147862	Homo sapiens	1.68e-25	More...	
E-GEOD-3189	Transcription profiling of human tissue samples to identify novel genes associated with malignant melanoma but not benign melanocytic lesions	malignant <i>melanoma</i> (diseasestate)	C1orf118	ENSG00000182705	Homo sapiens	2.64e-25	More...	

Terms of Use | [EBI Funding](#) | [Contact EBI](#) | © European Bioinformatics Institute 2006-2008. EBI is an Outstation of the [European Molecular Biology Laboratory](#).

Done

EFO in Atlas (www.ebi.ac.uk/microarray-as/atlas)

[home](#) | [about the project](#) | [faq](#) | [feedback](#) | [blog](#) | [help](#)



ATLAS
BETA

Genes: (all genes) Conditions: up/down in ▾ embryonic specific part Organism: Any species ▾

Search Atlas View results as: table heatmap expand conditions search with [EFO](#)

Your conditions query was expanded via [EFO](#) to: embryonic specific part "embryonic foregut sensory structure" "embryonic esophageal ganglion" "embryonic cuprophilic cell" "embryonic antennal sense organ" "embryonic foregut" "embryonic epipharynx" "embryonic esophagus" "embryonic corpus allatum" "embryonic central brain surface glia" "embryonic central brain glia" "embryonic dorsal apodeme" "embryonic corpus cardiacum" "embryonic anal pad" "embryonic brain" "embryonic gastric caecum" "embryonic dorsal epidermis" "embryonic central brain pars intercerebralis" embryo "embryonic central brain mushroom body" "embryonic frontal ganglion" "embryonic central brain neuron" "embryonic central brain" "embryonic ganglion mother cell" "embryonic central nervous system"

Displaying results 1-341 of ~341 found in the atlas

Experiment	Description	Factor Value (Factor)	Gene Name	Gene Id	Organism	P-value	AEW	...
E-AFMX-4	Transcription profiling of mouse cell types and tissues (GNF/Novartis)	tongue <i>epidermis</i> (organismpart)	Them5	ENSMJUSG00000028148	Mus musculus	1.87e-43	More...	
E-AFMX-4	Transcription profiling of mouse cell types and tissues (GNF/Novartis)	tongue <i>epidermis</i> (organismpart)	Krt13	ENSMJUSG00000044041	Mus musculus	3.87e-42	More...	
E-AFMX-4	Transcription profiling of mouse cell types and tissues (GNF/Novartis)	tongue <i>epidermis</i> (organismpart)	Krt4	ENSMJUSG00000059668	Mus musculus	1.26e-41	More...	
E-AFMX-4	Transcription profiling of mouse cell types and tissues (GNF/Novartis)	snout <i>epidermis</i> (organismpart)	Gpr115	ENSMJUSG00000023918	Mus musculus	7.11e-40	More...	
E-AFMX-4	Transcription profiling of mouse cell types and tissues (GNF/Novartis)	tongue <i>epidermis</i> (organismpart)	Serpib12	ENSMJUSG00000059956	Mus musculus	2.61e-37	More...	
E-AFMX-4	Transcription profiling of mouse cell types and tissues (GNF/Novartis)	vomerol nasal <i>organ</i> (organismpart)	Lcn3	ENSMJUSG00000028936	Mus musculus	3.44e-37	More...	

[Terms of Use](#) | [EBI Funding](#) | [Contact EBI](#) | © European Bioinformatics Institute 2006-2008. EBI is an Outstation of the [European Molecular Biology Laboratory](#).

Done

Challenges

- Multiple Inheritance: considered bad practice in ontology building – exists in many reference ontologies (and indeed life!)
- Upper Ontology integration
- Working with less mature ontology initiatives
- Release process for updating against external resources

Conclusions

- Present our application focused Experimental Factor Ontology
- Built with our needs in mind, however covers a wide range of experimental variables across a wide range of high-throughput technologies
- Reference to existing ontology resources when possible
- 0.2 release currently available in OLS
 - <http://www.ebi.ac.uk/ontology-lookup/browse.do?ontName=EFO>
 - <http://www.ebi.ac.uk/microarray-srv/efo/>

Acknowledgments

- **Ontology Creation:**
 - James Malone, Helen Parkinson
- **Mapping Tool and evaluation:**
 - Tim Rayner, Holly Zheng
- **External Review:**
 - Trish Whetzel, Jonathan Bard
- **Production Team:**
 - Anna Farne, Ele Holloway, Margus Lukk, Eleanor Williams, Tony Burdett
- Alvis Brazma, Misha Kapushesky
- Rebholz Group (Whatizit tool)