

Mouse Phenotyping Informatics Infrastructure (MPI2)

KOMP2 Project Kickoff

Paul Flicek

European Bioinformatics Institute

On behalf of

Bill Skarnes, Vivek Iyer, Dave Melvin (Wellcome Trust Sanger
Institute) Ann-Marie Mallon, Steve Brown (MRC Harwell)
Helen Parkinson, Gautier Koscielny, Terry Meehan* (EBI)

Overview

- Overall goal: Develop and deploy the IT infrastructure, database and web portal required to efficiently capture, manage, annotate, integrate and disseminate the phenotyping data from the KOMP2 program to the scientific and biomedical communities in an accurate, timely and intuitive manner.

Overview

- MPI2 partners have complementary skills and a record of support and delivery for large-scale genetics and genomics projects
 - Extensive experience in mouse mutagenesis, mouse phenotyping and mouse informatics
 - Committed to data release, academic and industry engagement
 - Assured archival, service delivery and computational infrastructure
 - Extensive collaborative links between the applicants and with worldwide partners

EBI: sequence to systems biology

- Core Mission: data integration, curation, dissemination, archiving
 - 3 million unique visitors per year
- 50% external grant funding from EC, Wellcome Trust, NIH etc
- Economies of scale, ~500 personnel, 50+ databases with scope: project specific to international collaboration
- Mouse Informatics Team
 - Current staff: 2x web developers, database developer, project lead
 - Supports EMMA, I-DCC, CREATE, EUCOMMTOOLS, INFRAFRONTIER, INFRAKOMP
- Expertise in managing high throughput project data – 1000 genomes, IHEC
- Expertise in content curation, annotation, collaborative ontology development
- Long term data archiving (since 1985)
- Dedicated systems support: 8000 core compute farm, 14 petabytes disk, managed servers, 3 off site data centres

MPI2 Aims

- **Aim 1:** Develop and deploy the Data Coordination Center (DCC)
- **Aim 2:** Perform continuous data annotation and statistical analysis
- **Aim 3:** Create a data center and IT infrastructure
- **Aim 4:** Develop and deploy the KOMP2 (IMPC?) web portal
(www.mousephenotype.org)

MPI2 Components

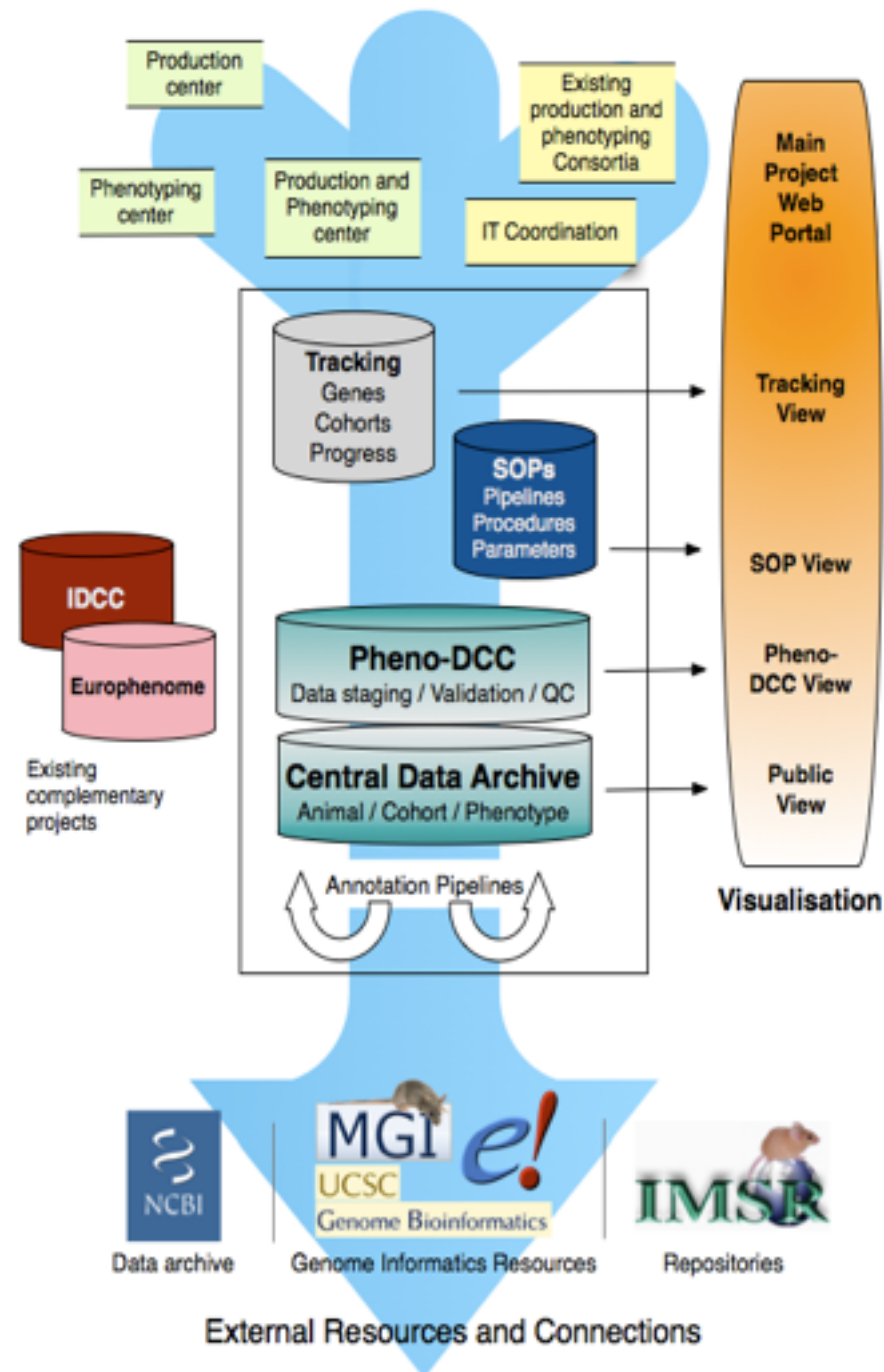
- **1. DCC**
 - KOMP2 data tracking
 - SOP data management (SOPDB)
 - Data upload, validation and quality control
 - (Pheno-DCC) Image data upload and annotation
- **2. Statistical analysis and data annotation**
- **3. Data center and IT infrastructure**
 - Core Data Archive (CDA)
 - Data access
 - Data integration with other informatics resources

MPI2 Components

- **4. KOMP2 web portal**
 - Data search
 - Data analysis and visualization
 - Data mining
 - Visualization of phenotypic variation and image data
 - User focused statistical analysis
 - Integrated data queries and semantic clustering
- **5. Community Outreach**


Component	Lead	Partners
Project Management, Reporting and Community Outreach	EBI	Harwell/Sanger
Tracking Database	Sanger	Harwell
SOP and Pheno-DCC	Harwell	Sanger
Annotation Pipelines	Harwell	Sanger/EBI
Central Data Archive	EBI	Sanger/Harwell
Web Portal	EBI	Sanger/Harwell

- Key existing technologies
 - EMPReSS; Europhenome
 - I-DCC, iMITs
 - Atlas, Ensembl, R-cloud



iMits – production planning and tracking

www.mousephenotype.org/imits





You are logged in as [wl@sanger.ac.uk](#)
[Change password](#) | [Logout](#)


[Home](#) | [Gene Selection](#) | [Search & Edit MI Attempts](#) | [Create MI Attempt](#) | [Reports](#)


Welcome to iMits


What would you like to do today?


**Gene Selection / Micro-Injection Planning**
Register interest (on behalf of your consortium) in micro-injecting some genes

**Search For & Edit Micro-Injection Attempts**
Search for an edit data existing micro-injection attempt records

**Create Micro-Injection Attempts**
Enter data on new micro-injection attempts

**Reporting**
View reports on all data in iMits


**Consortia**
View a list of all consortia submitting data to iMits and the production centres that are working as part of these consortia

**Users/Production Centres**
View a list of all iMits users and the production centre that they belong to

V. Iyer; WTSI

iMits – production planning

www.mousephenotype.org/imits



You are logged in as [wl@sanger.ac.uk](#)
[Change password](#) | [Logout](#)

HomeGene SelectionSearch & Edit MI AttemptsCreate MI AttemptReports

Register Interest in Micro-Injecting Genes

► [Filter Genes](#)

Please Select the Genes You Would Like to Register Interest In

Consortium: Production Centre: Priority: [+ Register Interest](#)

<input type="checkbox"/>	Gene	# IKMC Projects	# Clones	Non-Assigned MIs	Assigned MIs	Aborted MIs	MIs in Progress	GLT Mice
<input type="checkbox"/>	061004001Rik	2						
<input type="checkbox"/>	0610043K17Rik							
<input type="checkbox"/>	0910001L09Rik	2						
<input type="checkbox"/>	1010001B22Rik							
<input type="checkbox"/>	1010001N08Rik							
<input type="checkbox"/>	1100001G20Rik	3	11 Conditional6 Tar...				[DTCC-KOMP:...	

Page 2 of 2126 [Displaying 21 - 40 of 42505](#)

V. Iyer; WTSI

iMits – production planning

www.mousephenotype.org/imits



You are logged in as [wl@sanger.ac.uk](#)
[Change password](#) | [Logout](#)

[Home](#) | [Gene Selection](#) | [Search & Edit MI Attempts](#) | [Create MI Attempt](#) | [Reports](#)


Conflicting Micro-Injection Plans

Consortium	Production Centre	Marker Symbol	MGI Accession ID	Priority	Reason for Conflict
BaSH		Cyp2b10	MGI:88598	Medium	Other MI plans for: Helmholtz GMC
BaSH		F2rl1	MGI:101910	High	Other MI plans for: Phenomin
BaSH		Flnb	MGI:2446089	Low	Other MI plans for: JAX
BaSH	BCM	Lss	MGI:1336155	High	Other MI plans for: Helmholtz GMC
BaSH		Map3k14	MGI:1858204	Low	Other MI plans for: MRC
BaSH	BCM	Mapk4	MGI:2444559	Medium	Other MI plans for: Monterotondo

V. Iyer; WTSI

iMits – production tracking

www.mousephenotype.org/imits



You are logged in as [vvi@sanger.ac.uk](#)
[Change password](#) | [Logout](#)

[Home](#) | [Gene Selection](#) | [Search & Edit MI Attempts](#) | [Create MI Attempt](#) | [Reports](#)

Search for Micro-Injections

Micro-Injection Attempts

Everything	Transfer Details	Litter Details	Chimera Mating Details	QC Details	Summary						
ID	Edit In Form	ES Cell	Marker Symbol	Allele symbol	MI Date	Status	Colony Name	Consortium	Production Centre	Distribution Centre	Depos
100	Edit in Form	EPD0023_2_A12	Tcf7	Tcf7 ^{tm1a} (EUCCOM...	17-04-2008	Genotype confirmed	ICS-EPD0023_2...	EUCCOM-EUMO...	ICS	ICS	Froz
101	Edit in Form	EPD0038_2_B10	Xbp1	Xbp1 ^{tm1a} (EUCCOM...	10-10-2007	Micro-injection aborted	ICS-EPD0038_2...	EUCCOM-EUMO...	ICS	ICS	Froz
102	Edit in Form	EPD0021_1_F09	Prpsap2	Prpsap2 ^{tm1a} (EU...	10-10-2007	Genotype confirmed	ICS-EPD0021_1...	EUCCOM-EUMO...	ICS	ICS	Froz
103	Edit in Form	EPD0039_1_C07	Srsf4	Srsf4 ^{tm1a} (EUCCOM...	05-11-2007	Genotype confirmed	ICS-EPD0039_1...	EUCCOM-EUMO...	ICS	ICS	Froz
104	Edit in Form	EPD0038_2_E10	Xbp1	Xbp1 ^{tm1a} (EUCCOM...	07-11-2007	Genotype confirmed	ICS-EPD0038_2...	EUCCOM-EUMO...	ICS	ICS	Froz
105	Edit in Form	EPD0038_4_D01	Ube2b	Ube2b ^{tm1a} (EUCCOM...	07-11-2007	Genotype confirmed	ICS-EPD0038_4...	EUCCOM-EUMO...	ICS	ICS	Froz
106	Edit in Form	EPD0040_1_B07	Ulk2	Ulk2 ^{tm1a} (EUCCOM...	12-11-2007	Genotype confirmed	ICS-EPD0040_1...	EUCCOM-EUMO...	ICS	ICS	Froz
107	Edit in Form	EPD0039_3_C02	Grih3	Grih3 ^{tm1a} (EUCCOM...	12-11-2007	Genotype confirmed	ICS-EPD0039_3...	EUCCOM-EUMO...	ICS	ICS	Froz
108	Edit in Form	EPD0038_2_B10	Xbp1	Xbp1 ^{tm1a} (EUCCOM...	17-01-2008	Micro-injection aborted	ICS-EPD0038_2...	EUCCOM-EUMO...	ICS	ICS	Froz
109	Edit in Form	EPD0061_1_C10	Fabp3	Fabp3 ^{tm1a} (EUCCOM...	18-02-2008	Genotype confirmed	ICS-EPD0061_1...	EUCCOM-EUMO...	ICS	ICS	Froz
110	Edit in Form	EPD0059_4_D03	Ngfr	Ngfr ^{tm1a} (EUCCOM...	22-02-2008	Genotype confirmed	ICS-EPD0059_4...	EUCCOM-EUMO...	ICS	ICS	Froz
111	Edit in Form	EPD0065_5_A04	Jmjd5	Jmjd5 ^{tm1a} (EUCCOM...	25-02-2008	Genotype confirmed	ICS-EPD0065_5...	EUCCOM-EUMO...	ICS	ICS	Froz

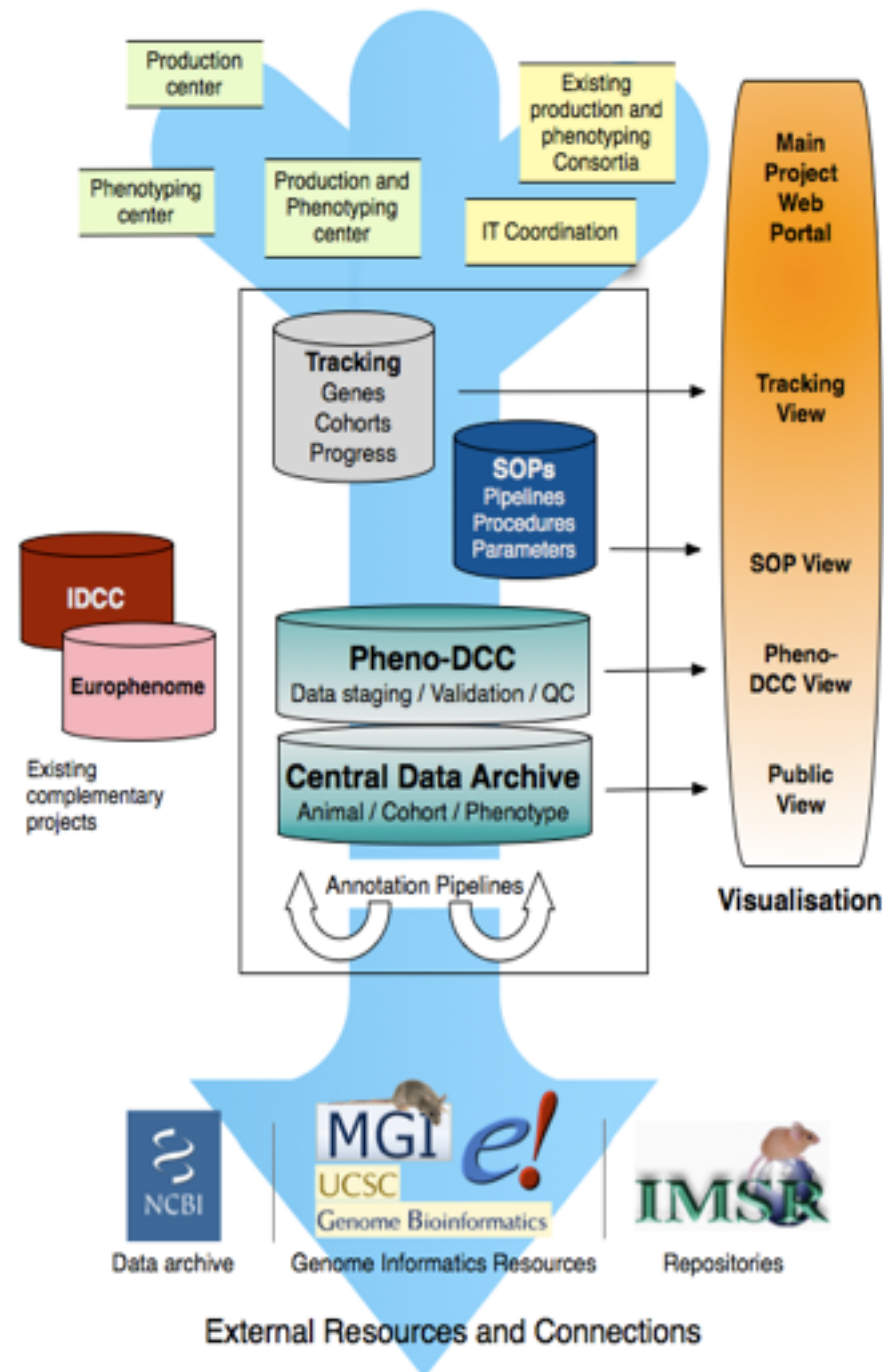
Page 1 of 10

Displaying 1 - 20 of 197

V. Iyer; WTSI

Component	Lead	Partners
Project Management, Reporting and Community Outreach	EBI	Harwell/Sanger
Tracking Database	Sanger	Harwell
SOP and Pheno-DCC	Harwell	Sanger
Annotation Pipelines	Harwell	Sanger/EBI
Central Data Archive	EBI	Sanger/Harwell
Web Portal	EBI	Sanger/Harwell

- Key existing technologies
 - EMPReSS; Europhenome
 - I-DCC, iMITs
 - Atlas, Ensembl, R-cloud



EuroPhenome (www.europhenome.org)

- Data Coordination Centre and web portal for EUMODI C project.
- Captures data from;
 - Wellcome Trust Sanger Institute
 - , UK MRC Harwell, UK
 - Helmholtz Zentrum Munchen, Germany
 - Institut Clinique de la Souris, France.
 - Centre for Modeling Human Disease, Canada



SOPs Definition



Defining data Standards



Data Capture in LIMS



Data Export to Centralised DB



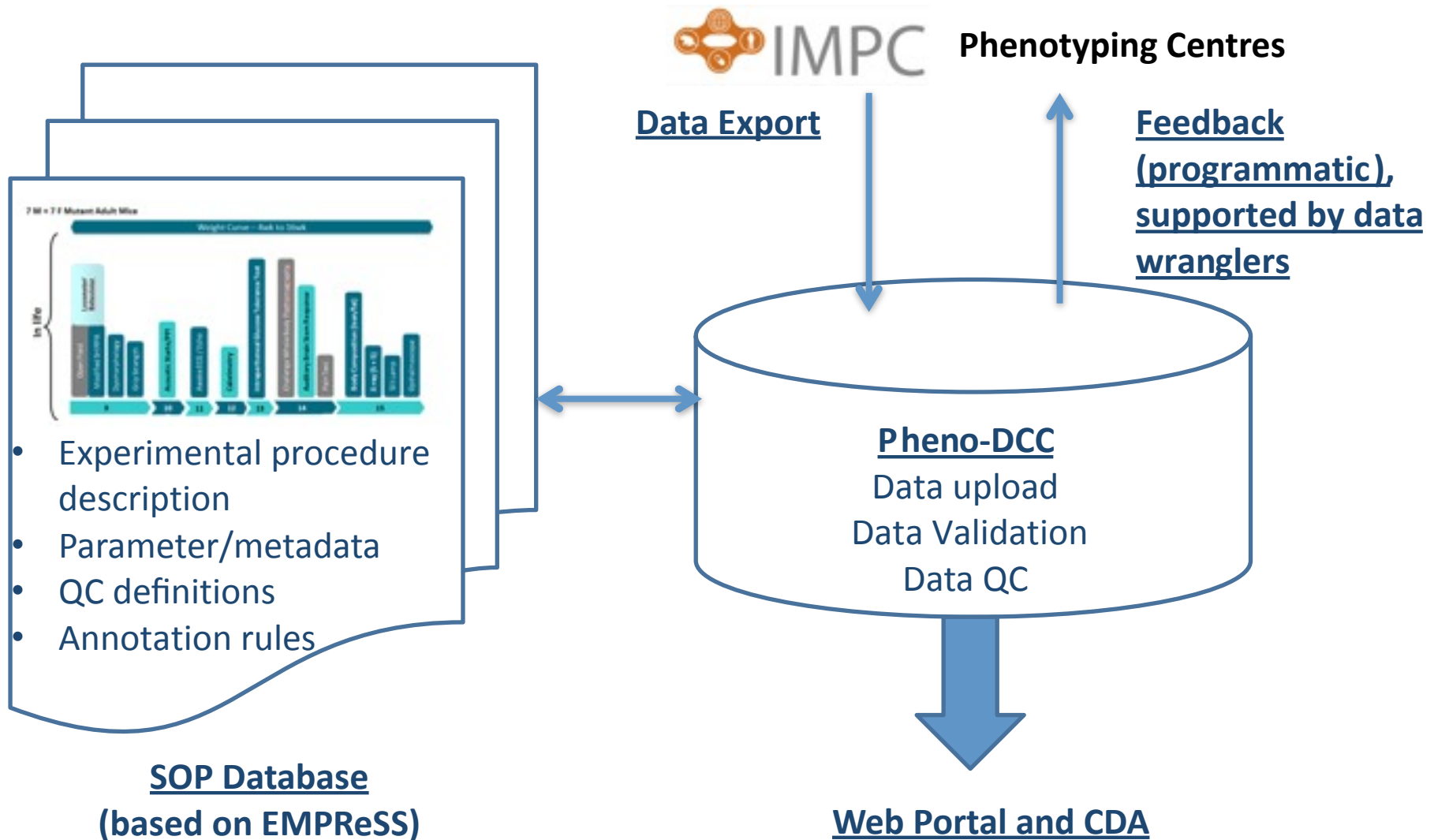
Validation and QC



Data Capture Pipeline

A. Mallon; Harwell

Data Coordination Centre (Pheno-DCC)



A. Mallon; Harwell

KOMP2 SOP Database

- Review all existing SOPs, collaborate with phenotyping centres to define and standardise new SOPs in KOMP2/IMPC pipeline (Data Wranglers)
- Add further data to SOP to aid QC and validation (e.g. data bounds); collaboration with BioStatistician and Data Wranglers
- Build on EMPReSS database, to develop new functionalities including SOP management and version control
- Extend portal to have levels of user access (read/write)
 - NIH Staff; Phenotyping Centres; Scientific Community; Public

Main
Project
Web
Portal

Tracking
View

SOP View

Pheno-
DCC View

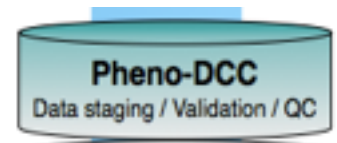
Public
View

SOPs
Pipelines
Procedures
Parameters

A. Mallon; Harwell

Pheno-DCC Data Export and Upload Process

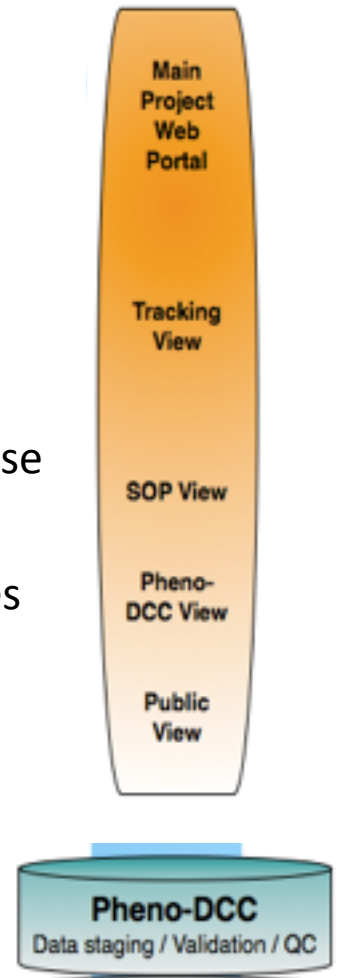
- Data Wranglers work with centres to review data export and upload processes to ensure compatibility
- Design and review XML schemas
- Analyse and document roles, ownership and responsibilities relating to data submission
- Design and implement an export module for plugin to different LIMS
- Support installation of this module at centres (Data Wranglers)
- Implement web based upload and editing



A. Mallon; Harwell

Pheno-DCC Data Validation and Quality Control

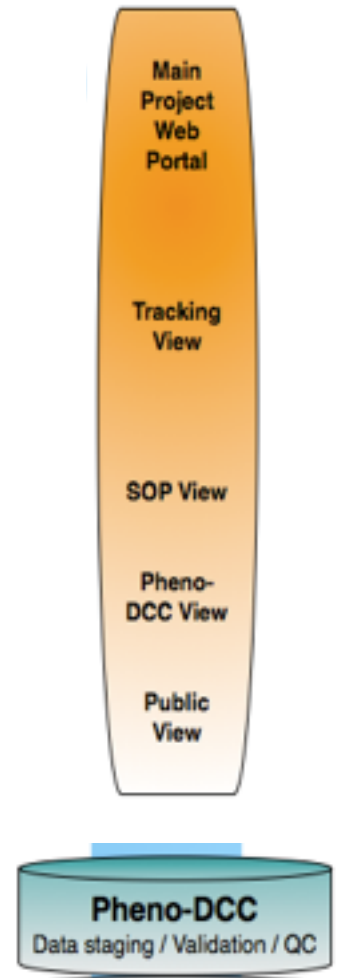
- Data Validation
 - Develop and expand tools to reject invalid data at point of submission
 - Ensure inclusion of valid minimum information
 - Investigate causes of missing data
 - Document and track changes made by Data Wranglers, report these via a web portal
 - Develop and test automated tools to take over some of these roles



A. Mallon; Harwell

Data Validation and Quality Control

- Data Quality Control
 - Develop QC supporting tools to flag potential data issues
 - Investigate flagged issues
 - Particular emphasis on:
 - Systematic drift within a centre
 - Operator bias
 - Differences between centres
 - Unexpectedly high rates of change
- Develop a portal as part of mousephenotype.org where consortium members, NIH staff can view progress of submission, data errors etc



A. Mallon; Harwell

What is a data wrangler?

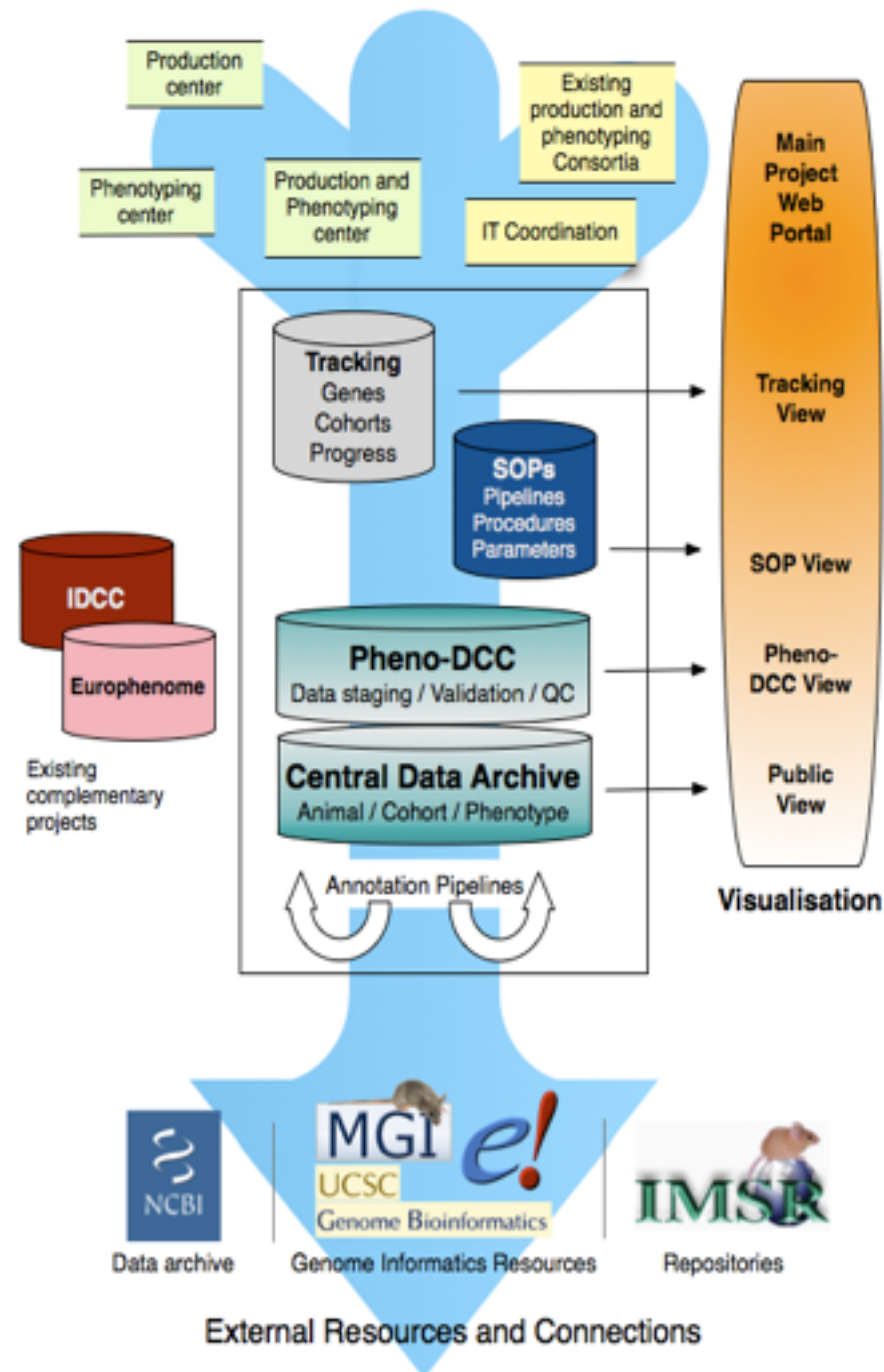
- SOP
 - Centralised standardisation
 - Parameters / Derived Parameters
 - Metadata
 - Graph Type; Statistical Analysis
 - Ontological Annotations
 - Collaboration with phenotyping centres to setup the SOPS in LIMS

What is a data wrangler?

- Animal and Phenotyping Data
 - Timely submission of data; managing edits/deletes
 - Erroneous values
 - Decimal place errors; Missing values; Outliers; Categorical Parameters (N, No, no) etc
 - Submission of data for which we have no mice
 - Centralised QC
 - Identify large variations in data across centres
 - Baseline strain data analysis; drift, equipment issues, experimenter issues, annotation issues etc
 - Curation
 - IDs for Allele are correct with data in MGI

Component	Lead	Partners
Project Management, Reporting and Community Outreach	EBI	Harwell/Sanger
Tracking Database	Sanger	Harwell
SOP and Pheno-DCC	Harwell	Sanger
Annotation Pipelines	Harwell	Sanger/EBI
Central Data Archive	EBI	Sanger/Harwell
Web Portal	EBI	Sanger/Harwell

- Key existing technologies
 - EMPReSS; Europhenome
 - I-DCC, iMITs
 - Atlas, Ensembl, R-cloud



KOMP2 Statistical Analysis Overview

- Early phase re-use of Harwell's bespoke Europhenome pipeline
- Development of new methods BioStatistican WTSI)
- Port redeveloped pipeline to EBI R environment delivering scaling, transparency and modularity <http://www.ebi.ac.uk/Tools/rcloud/>
- Collaborative R workbench for project and external use running on EBI resources
- Fast access via Core Data Archive, R, API access
- GUI integration year 3
 - Use case: identify data, define your R analysis and run remotely
 - Use case: repeat an existing project defined R analysis on subsets of project data
 - Use case: Run existing pipeline on revised project data
 - Use case: Store project pipelines with provenance
 - Use case: Store user defined pipeline with provenance



KOMP2 Statistical Analysis and Data Annotation

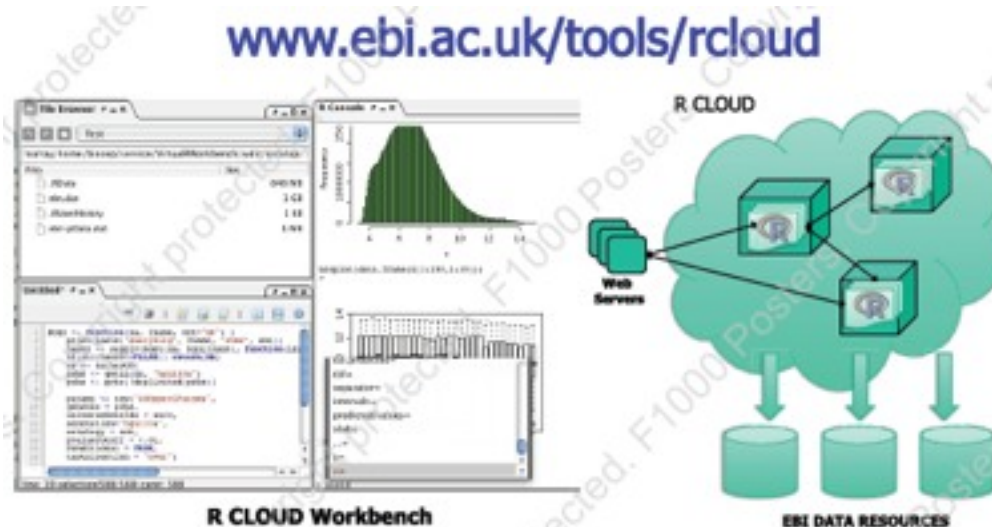
- Extend, redevelop and scale existing automated annotation pipeline (AAP) to assign phenotype ontology terms to statistically significant phenodeviant
- Review experimental design and statistical data analysis for each procedure and alter AAP accordingly
- Coordinate with data wranglers and domain experts to capture definition of ontology terms related to each parameter
- Incorporate image based procedures, initially LacZ, into AAP including EBI R analysis infrastructure

KOMP2 Statistical Analysis and Data Annotation

- Incorporate data from existing mouse gene function databases and human GWAS projects to build a picture of phenotypic variability
- Develop use cases to leverage the complete set of integrated annotation
- Make the statistical tools publicly available through the web portal, and allow expert users to define their own methods of data analysis

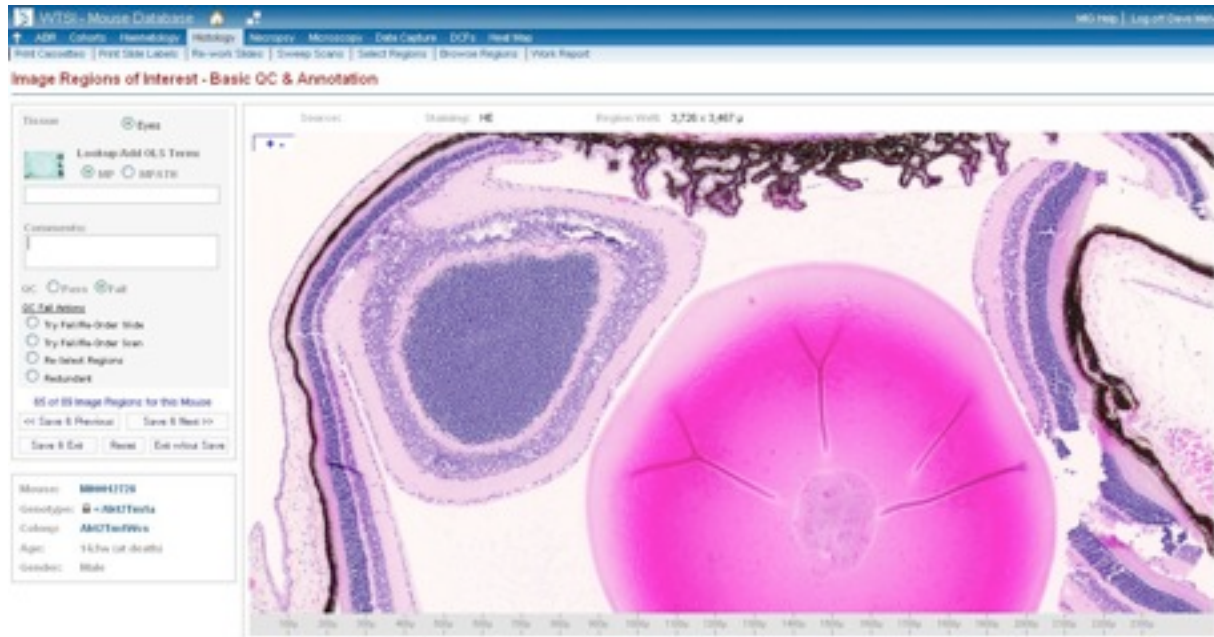
Analysis infrastructure

- R environment
 - Bioconductor package authors and users
 - Bransparent power user access via R wiki
 - Modular plug and play statistical pipeline development
 - R cluster available (for those really big analyses)
 - R workbench
 - Access to public data without download



H. Parkinson; EBI

Imaging (a future problem)

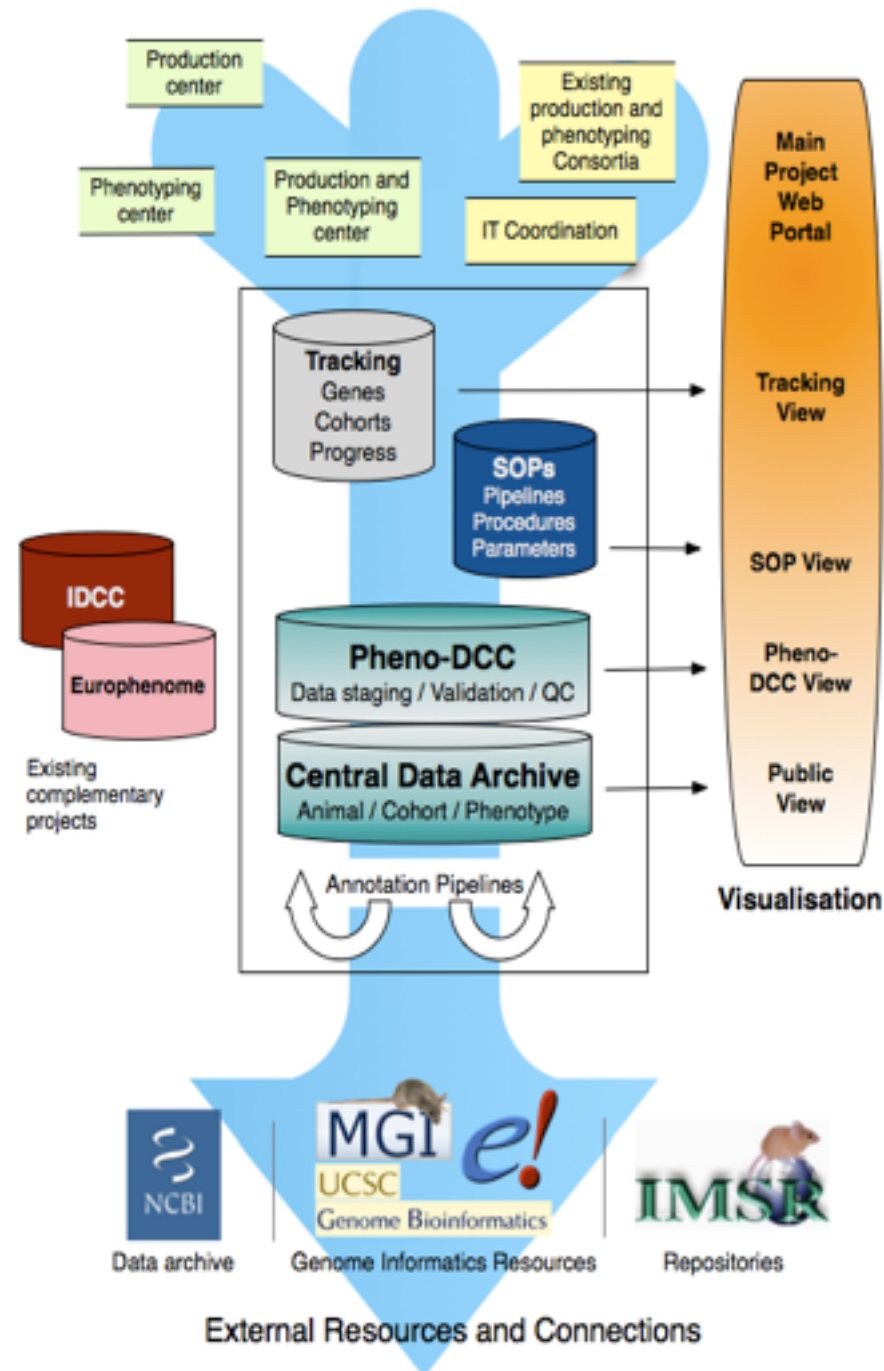


- Imaging
 - The 1 FTE available from KOMP2 will only be able to deliver modest objectives.
 - If planned/hoped advances are realised during the project will approach appropriate funding agencies with a proposal to provide additional, complementary, imaging work for example 3D MRI and micro-CT

D. Melvin; WTSI

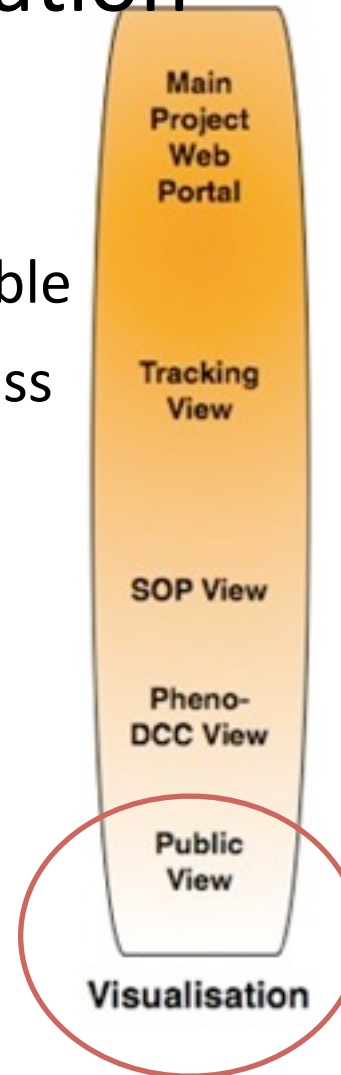
Component	Lead	Partners
Project Management, Reporting and Community Outreach	EBI	Harwell/Sanger
Tracking Database	Sanger	Harwell
SOP and Pheno-DCC	Harwell	Sanger
Annotation Pipelines	Harwell	Sanger/EBI
Central Data Archive	EBI	Sanger/Harwell
Web Portal	EBI	Sanger/Harwell

- Key existing technologies
 - EMPReSS; Europhenome
 - I-DCC, iMITs
 - Atlas, Ensembl, R-cloud



Core Data Archive (CDA) implementation

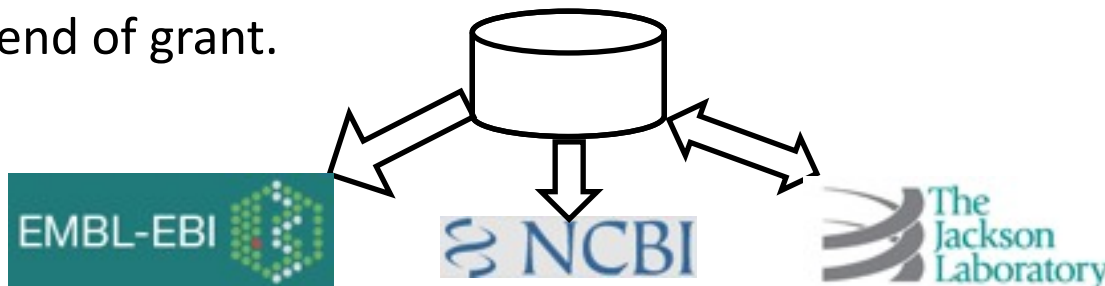
- Project specific database
- Data is accessioned, versioned, updateable, reanalysable
- Centralized for robustness, scalable external user access
- Complex update model – flexibility
- Multiple data access/export modes
 - Supplier of CDA data for web portal
 - DAS access for S4 integration
 - APIs, R, web services
 - File push to NCBI/Jackson labs
 - Project FTP site



H. Parkinson; EBI

Data archiving and exchange

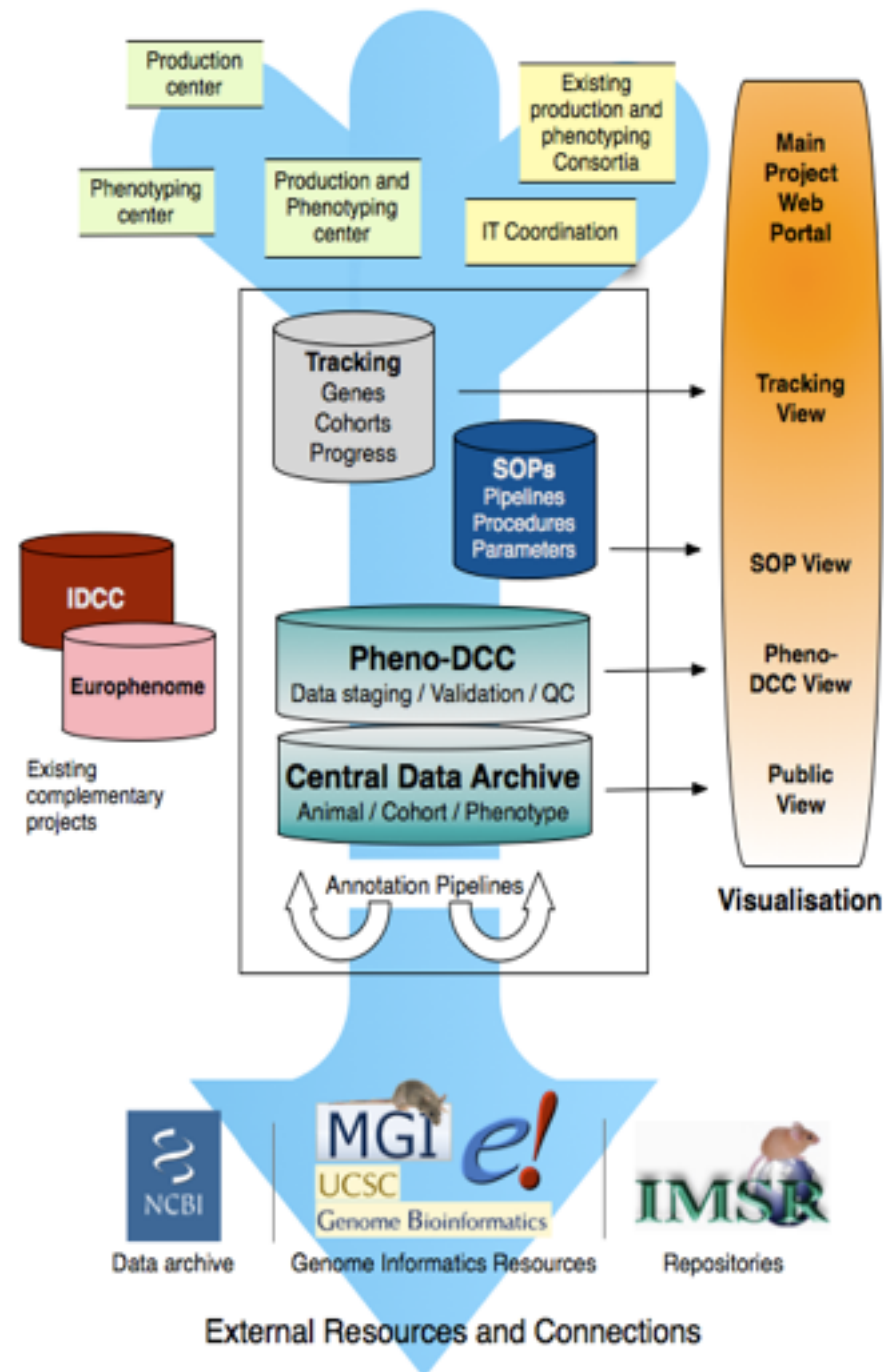
- Pipeline for direct deposition to NCBI/EBI from the DCC
 - dbGAP, BioSamples and other appropriate databases
- Periodic release process for QCed data
 - Agreed with phenotyping centres/NIH
- Identify and implement an appropriate update model
 - Corrections, data withdrawal etc
- Work with the KOMP2 users and archive to ensure appropriate data is archived for future usability - this will determine the content of archives e.g. raw and summary level data.
- Work with the NCBI/EBI to explore image archiving needs as data appears
- During year 5 review all archived data and ensure these are updated by end of grant.



P. Flicek/H. Parkinson; EBI

Component	Lead	Partners
Project Management, Reporting and Community Outreach	EBI	Harwell/Sanger
Tracking Database	Sanger	Harwell
SOP and Pheno-DCC	Harwell	Sanger
Annotation Pipelines	Harwell	Sanger/EBI
Central Data Archive	EBI	Sanger/Harwell
Web Portal	EBI	Sanger/Harwell

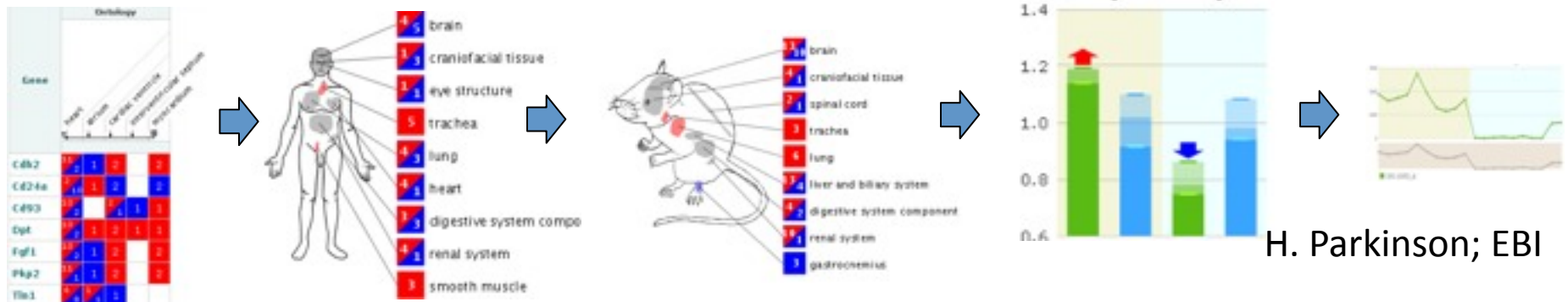
- Key existing technologies
 - EMPReSS; Europhenome
 - I-DCC, iMITs
 - Atlas, Ensembl, R-cloud



Existing Web portals

- Atlas

- Generic data warehouse (gene expression, proteomics, annotation database, R backend), gene annotation updates from Ensembl/other sources
- Sample annotations, ontology support, complex queries
- Distributed, open source
- Combined semantic and statistical queries
- Complex GUI, clickable images, heat maps
- Summary level data presentation, access to statistical pipeline, multiple plots, p values, alternative views, customizable
- Similarity searching by expression pattern



H. Parkinson; EBI

Gene search

Find Gene:

eg. [Akt2](#)
[+ Advanced Search Options](#)

Phenotype search

Find MP Term:

eg. [abnormal glucose homeostasis](#)
[+ Advanced Search Options](#)

Summary : [Crtc1](#) (CREB regulated transcription coactivator 1)

Genotype Information:

Gene Symbol [Crtc1](#)
Gene name CREB regulated transcription coactivator 1
International strain name C57BL/6NTac-Crtctm1a(EUCOMM)Wts/ncs
Allele Symbol [Crtc1^{tm1a\(EUCOMM\)Wts/ncs}](#)
Ensembl Gene Id [ENSMUSEG00000031575](#)
EMMA [EM 04392](#)

Assoc. Human Disorders [Mucoepidermoid salivary gland carcinoma](#)

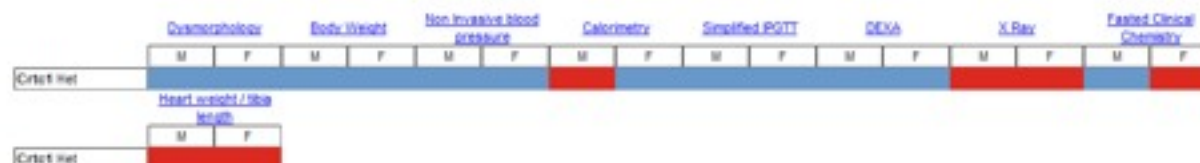
Contact [EUMODIC](#) for further information

Pipelines :

Pipeline	Number of annotated parameters
EUMODIC Pipeline 1	6
EUMODIC Pipeline 2	6

[Homocysteine viability at weaning](#)
Outcome Both Sexes Homozygous - Lethal
[Reporting the fertility of homozygous GA mouse lines](#)
Outcome Both Not applicable

EUMODIC Pipeline 1 (Click to select)



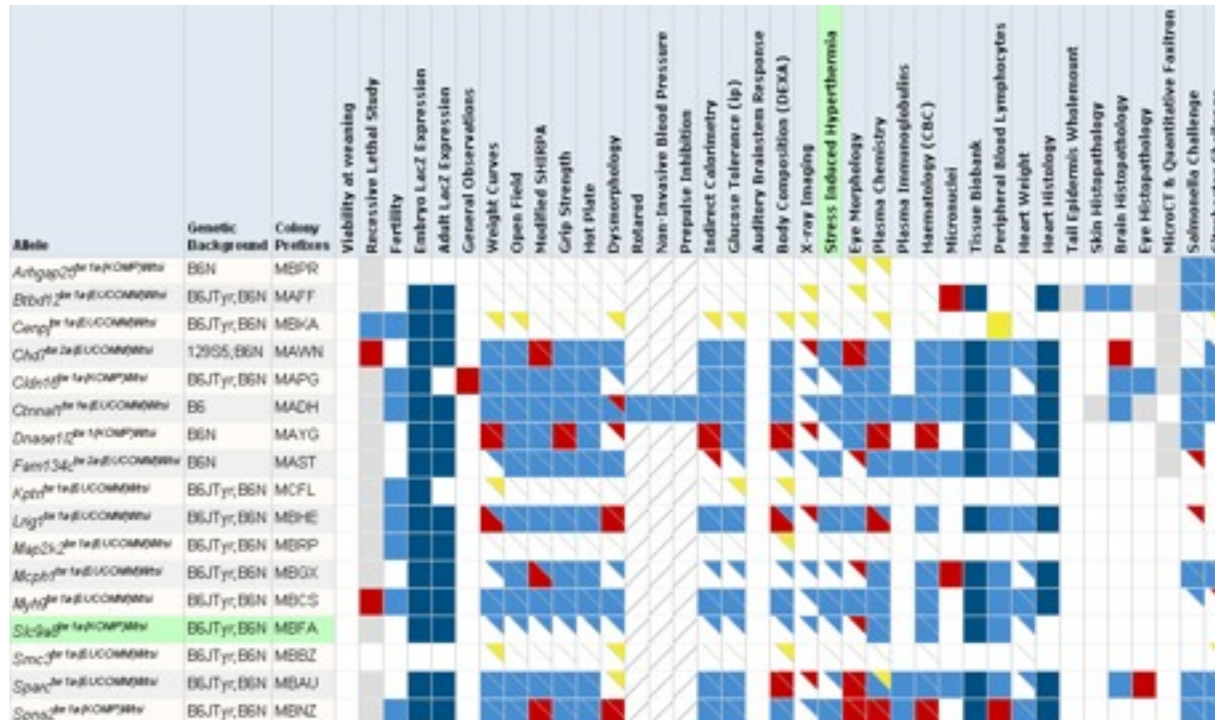
Legend: No Data No Significant Annotations Significant Annotation Present

EUMODIC Pipeline 2 (Click to select)



A. Mallon; Harwell

Existing WTSI Data Visualisation

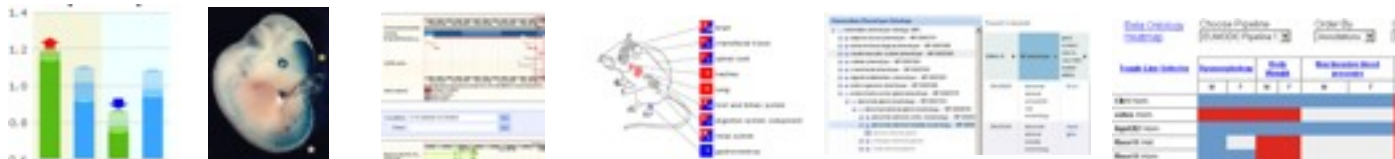


- Review primary visualisation technology
 - WTSI and Harwell approaches to heat-mapping as the starting point.
 - Establish the formal framework for critique of software tools.

D. Melvin; WTSI

Planned KOMP2 public web portal

- Single point of entry
- Collaborative development and delivery – many views and user groups internal and external
- Best of breed design building on prior experience from all partners
 - User focussed; Search and identification; Complex image rendering
 - Data mining – pheno-pheno/pheno-locus
 - User profiles, saved queries, watch genes, RSS
 - feeds Wizards for complex query design
 - Direct statistical analyses
 - Canned reports for common queries
 - Downloads
 - Ontology search and visualization
- Technology use matches the tasks and priorities – e.g. tracking is delivered before complex image visualization
- As data volume grows visualizations are modified



Existing data integration methods and added value analysis

- Integration with GWAS data via disease terminology mapping and locus
- Ensembl integration – locus mapping, genome annotation integrated views

e!Ensembl BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors

Mouse (NCBI M37) Location: 11:101,031,837-101,052,038 Gene: Cntnap1

Gene-based displays

- Gene summary
- Splice variants (2)
- Supporting evidence
- Sequence
- External references
- Regulation
- Comparative Genomics
 - Genomic alignments
 - Gene Tree (image)
 - Gene Tree (text)
 - Gene Tree (alignment)
 - Orthologues (3)
 - Paralogues (3)
 - Protein families (1)
- Genetic Variation
 - Variation Table
 - Variation Image
- External Data
 - Personal annotation
 - ID History
 - Gene history

Configure this page

Manage your data

Export data

Bookmark this page

Download view as CSV

Gene: Cntnap1 (ENSMUSG00000017167)

Description: contactin associated protein-like 1 Gene [Source:MGI Symbol;Acc:MGI:195600]

Location: [Chromosome 11: 101,031,837-101,052,038](#) forward strand.

Transcripts: There are 2 transcripts in this gene

Name	Transcript ID	Length (bp)	Protein ID	Length (aa)	Biotype	CCDS
Cntnap1-001	ENSMUST00000103109	5440	ENSMUSP00000099099	1365	Protein coding	CCDS25455
Cntnap1-002	ENSMUST00000138942	664	No protein product	-	Processed transcript	-

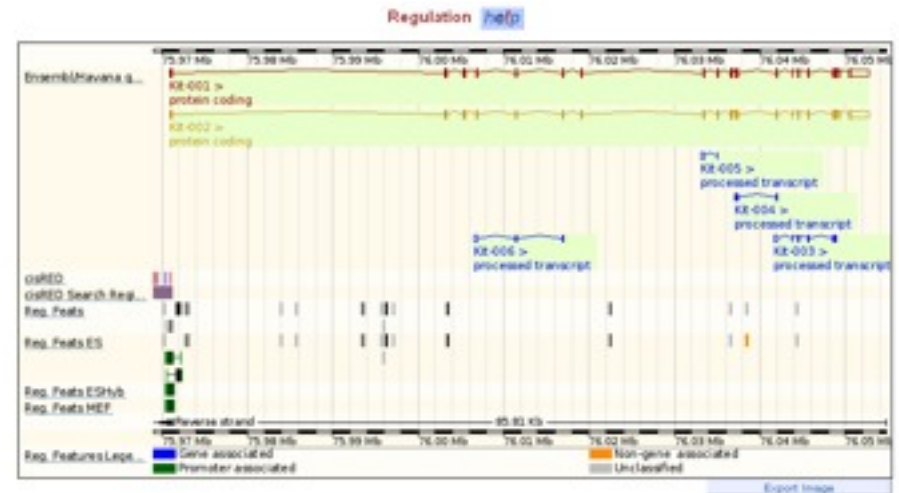
External references [help](#)

This gene corresponds to the following database identifiers:

Show/hide columns	Database identifier
International Knockout Mouse Consortium Knockouts (ES cells available)	Cntnap1 [view all locati
MGI Symbol	Cntnap1 contactin associated p

The following database identifiers correspond to Transcripts of this Gene:

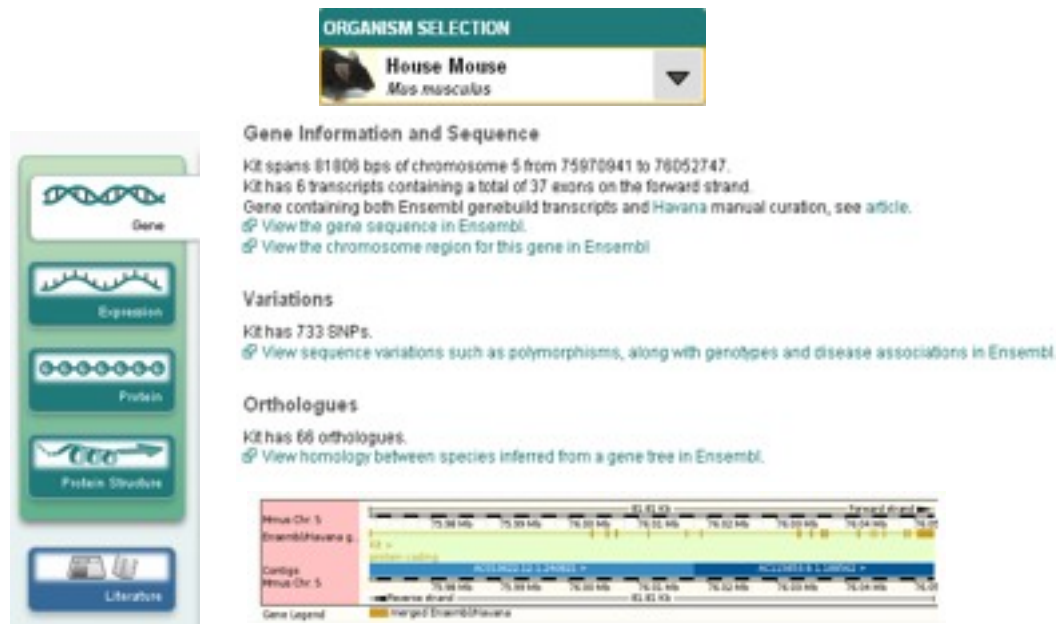
Transcript ID	MGI Symbol	EntrezGene	CCDS	UniProtKB/Swiss-Prot	RefSeq peptide
ENSMUST00000103109	Cntnap1	Cntnap1	CCDS25455.1	CNTNP1_MOUSE	NP_05805
ENSMUST00000138942					



P. Flicek; EBI

Planned data integration and added value

- Integration into EBI existing spines (DAS based)
- Development of new spines diseases, cell type, tissue, tools
- User focussed design with general and specific user groups
- Added value - terminology, literature searching, pathways etc (user defined)
- Reciprocal integration between KOMP2 web portal and EBI resources



ORGANISM SELECTION

House Mouse
Mus musculus

Gene



Expression



Protein



Protein Structure



Literature

Alkaline phosphatase, liver/bone/kidney Gene

[View in Ensembl](#)

Gene Information and Sequence

Alpl spans 54651 bps of chromosome 4 from 137297648 to 137352298.

Alpl has 8 transcripts containing a total of 26 exons on the reverse strand.

Gene containing both Ensembl genebuild transcripts and Havana manual curation, see [article](#).

[View the gene sequence in Ensembl.](#)

[View the chromosome region for this gene in Ensembl](#)

Variations

Alpl has 513 SNPs.

[View sequence variations such as polymorphisms, along with genotypes and disease associations in Ensembl.](#)

Orthologues

Alpl has 69 orthologues.

[View homology between species inferred from a gene tree in Ensembl.](#)

Paralogues

Alpl has 3 paralogues.

[View homology arising from a duplication event, inferred from a gene tree in Ensembl.](#)

Regulation

There are 10 regulatory elements located in the region of Alpl.

[View the gene regulatory elements, such as promoters, transcription binding sites, and enhancers in Ensembl.](#)

ORGANISM SELECTION



House Mouse
Mus musculus



Gene



Expression



Protein



Protein Structure



Literature

Alkaline phosphatase, liver/bone/kidney Gene

All Articles (96)



In vivo molecular evidence of delayed titanium implant osseointegration in compromised bone.
Vandamme K, Holy X, Bensidhoum M, Logeart-Avramoglou D, Naert IE, Duyck JA and Petite H
Biomaterials 32(14) May 2011

Molecular characterization of tissue-nonspecific alkaline phosphatase with an Ala to Thr substitution at position 116 associated with dominantly inherited hypophosphatasia.
Ishida Y, Komaru K and Oda K
Biochimica et biophysica acta 1812(3) Mar 2011

Hypophosphatasia now draws more attention of both clinicians and researchers: A Commentary on prevalence of c. 1559delT in ALPL, a common mutation resulting in the perinatal lethal form of hypophosphatasias in Japanese and effects of the mutation on heterozygous carriers.
Ozono K and Michigami T
Journal of human genetics 56(3) Mar 2011

[View all 96 in CiteXplore](#)

Reviews (9)



Genetic aspects of osteoporosis.
Hosoi T
Journal of bone and mineral metabolism 28(6) Nov 2010

Diagnosis of bone metastases in urological malignancies—an update.
Rajarubendra N, Bolton D and Lawrentschuk N
Urology 76(4) Oct 2010

Hypophosphatasia.
Momet E
Best practice & research. Clinical rheumatology 22(1) Mar 2008

[View all 9 in CiteXplore](#)

Title Contains "Alpl" (15)



Hypophosphatasia now draws more attention of both clinicians and researchers: A Commentary on prevalence of c. 1559delT in ALPL, a common mutation resulting in the perinatal lethal form of hypophosphatasias in Japanese and effects of the mutation on heterozygous carriers.
Ozono K and Michigami T

Free Full Text Articles (12)



Genome-wide association study of vitamin B6, vitamin B12, folate, and homocysteine blood concentrations.
Tanaka T, Scheet P, Giusti B, Bandinelli S, Piras MO, Usala O, Lai S, Mulis A, Corsi AM, Vestri A, Sofi F, Gori AM, Abbate R, Guralnik J, Singleton A, Abecasis GR, Colaneri P, Uda M, et al

ORGANISM SELECTION



House Mouse
Mus musculus



Gene



Expression



Protein



Protein Structure



Literature

Alkaline phosphatase, liver/bone/kidney Gene

Alpl differential expression summary

[View in Gene](#)

Organism part

liver, kidney, placenta, thymus

[View all](#)

Disease state

normal, (empty), Pb-A infected, myocardial infarction

[View all](#)

Cell type

embryonic stem cell, hematopoietic stem cell, T effector cell, T regulatory cell

[View all](#)

Cell line

N1E-115 wild_type, 67NR, Swiss5, Swiss8

[View all](#)

Compound treatment

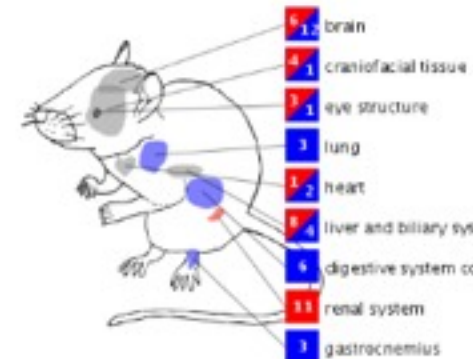
control, none, vehicle - arachis oil, 17beta-estradiol

[View all](#)

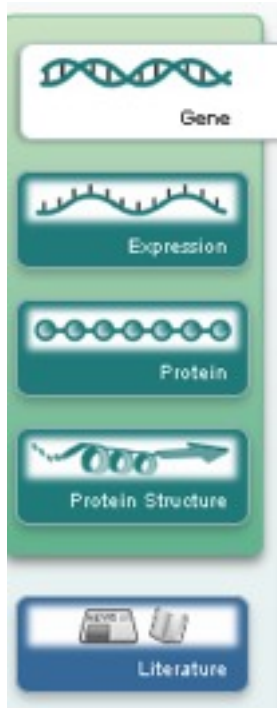
Developmental stage

embryo, adult, fetus, neonate

[View all](#)



[View](#) Number of published studies where the gene over/under compared to the gene's overall mean expression level



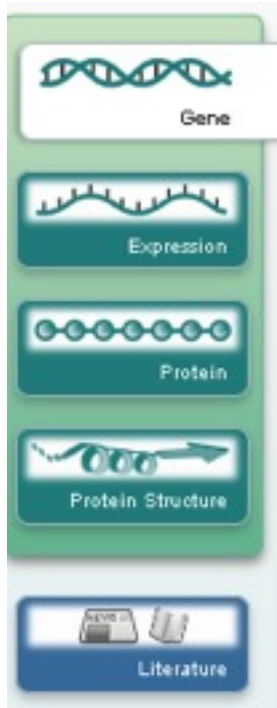
Disease

Pathways

Tissues

Chemistry

Tools



→ KOMP2 Ensembl links

→ LacZ summaries, image links



Disease

→ Mouse models of disease, phenotype summaries

Pathways

Tissues

→ Expression summaries, phenotype links

Chemistry

Tools

→ Mouse knockouts, phenotype summaries, CDA links

Existing Outreach Capacity

- Training and outreach team of 11, train on all EBI resources, 177 courses in 2010
- Focus on biological users in academia and industry (not bioinformaticians)
- Embedded experts in teams contribute content and delivery
- Targets academia, SME, big pharma
- Industry programme – research, training, collaboration

Astellas Pharma Inc.	AstraZeneca	Bayer Schering Pharma	Boehringer Ingelheim	Galderma
GlaxoSmithKline	EB Lilly and Company	F. Hoffmann-La Roche	Johnson & Johnson Pharmaceutical Research & Development	Merck Serono S.A.
Hestlé Research Centre	Novo Nordisk	Orion Pharma	Philips Research	Pfizer Ltd
Syngenta	Sanofi-Aventis	UCB	Unilever	

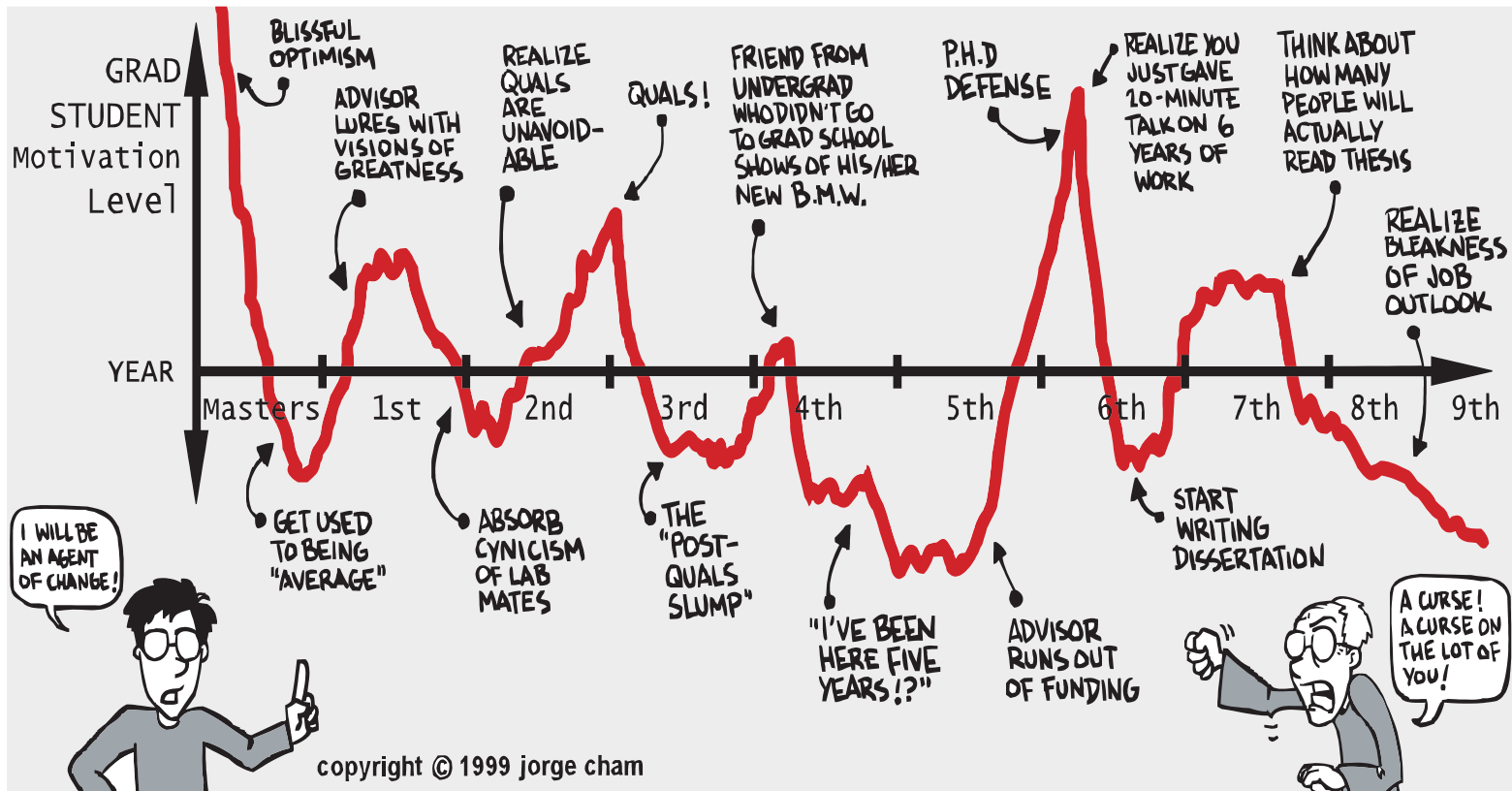
P. Flicek/H. Parkinson; EBI

WORKING WITH THE CONSORTIUM

09/28/11

Thursday, 29 September 2011

WHERE ARE YOU?



"Piled Higher and Deeper" by Jorge Cham

www.phdcomics.com

Four stages of a project DCC

- Stage 1: This can't be so hard
 - We'll just get a couple of undergraduates and some hard drives at Best Buy
 - It cost's how much?
 - You can only do what?
 - Danger is unrealistic expectatitions

Four stages of a project DCC

- Stage 2: The source of all problems in the project
 - These people are idiots and in the way
 - Why aren't they doing X, Y, Z and not A,B,C?
 - How come you didn't know that we had planned a press announcement for four months?
 - Perceived acceptable standard is perfection and only acceptable time frame is yesterday
 - This stage normally includes an “intervention”

Four stages of a project DCC

- Stage 3: It works! Mostly.
 - Lots of agreement on how things should be done
 - Mostly just getting on with it
 - The culmination of a tremendous amount of work
 - Best science done here
 - We will be supporting multiple constituencies
 - Production centers
 - KOMP2/IMPC affiliated analysis groups
 - Members of the community that want to use this data
 - NIH staff

Four stages of a project DCC

- Stage 4: See, that wasn't so hard
 - Characterised by unrealistic new requirements
 - “We just made an agreement to triple the size/complexity/usability of the data. Sorry we didn't tell you before, but this is ok right?”
 - “We going to extend the project by 2 years. There is no additional funding, but things are going pretty well for you, right?”
 - “This other group has a great idea for analysis/data flow. Can you quickly integrate all of their work?”

Getting though the stages

- Communication, communication, communication
 - We don't know all the details of what is going on either
 - We all have conflicting requirements and multiple top priority tasks
 - Help us though these as we all try to be understanding
- Pick up the phone:
 - Call me: +44 1223 492581
 - Each component has a lead PI
 - Harwell: Ann-Marie: +44 1235 84 1077
 - Sanger: Vivek: +44 1223 495 364
 - EBI: Helen (currently on maternity leave)
 - Set up a phone conference