Bio-Imaging for the International Mouse Phenotyping Consortium

5 June 2014

9th OME User's Meeting - Paris

Gautier Koscielny

MPI2 Consortium

EMBL-EBI

www.mousephenotype.org

IMPC

Outline

- Goals of IMPC
- Informatics for the IMPC
- Preliminary results
- 2D Imaging modalities (2 use cases)
- Work in progress: 3D Imaging modalities (2 use cases)
- Dissemination on the IMPC portal
- Future directions

IMPC

MPI2

IMPC

International Mouse Phenotyping Consortium

www.mousephenotype.org

From Steve Brown, MRC Harwell

IMPC Goals

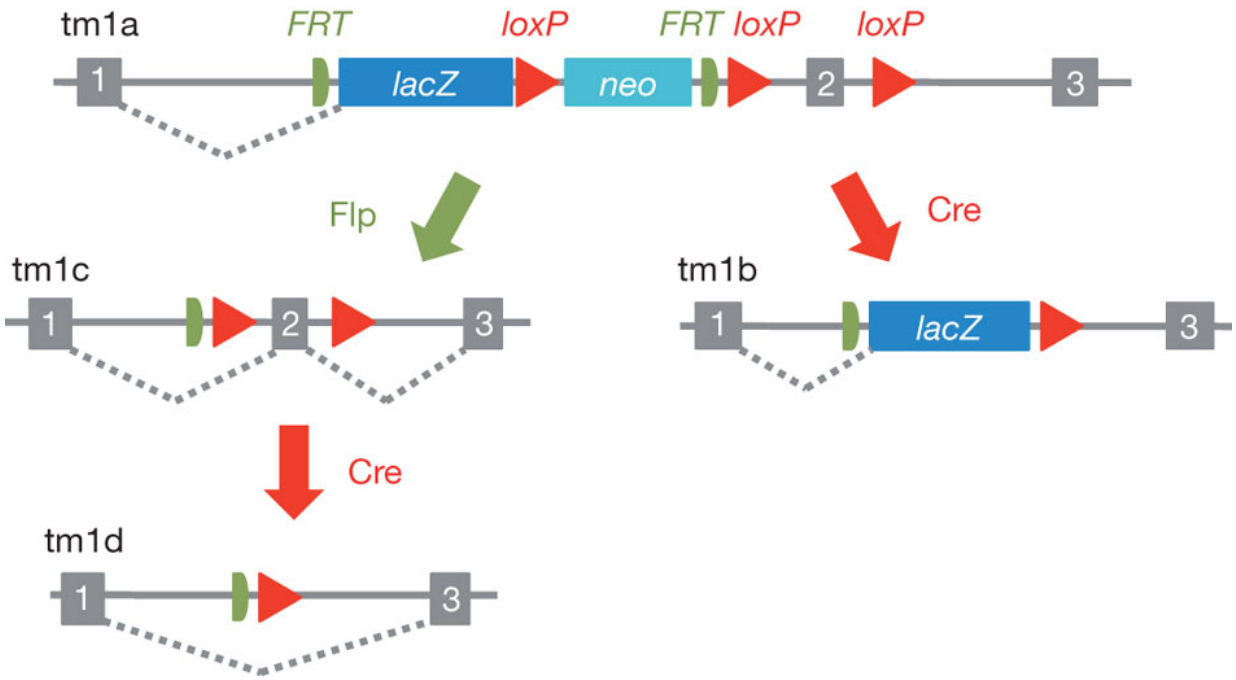
- Encyclopedia of mammalian gene function
- Create and phenotype over 20,000 mutant mouse lines
- Build collaborative "networks" for more focused phenotyping efforts
- Industry outreach
- Provide a centralized data center and portal for free, unrestricted access to primary and secondary data

IMPC

MPI2

Alleles most frequently used in IMPC

Schematic of the ‘knockout-first’ conditional allele.



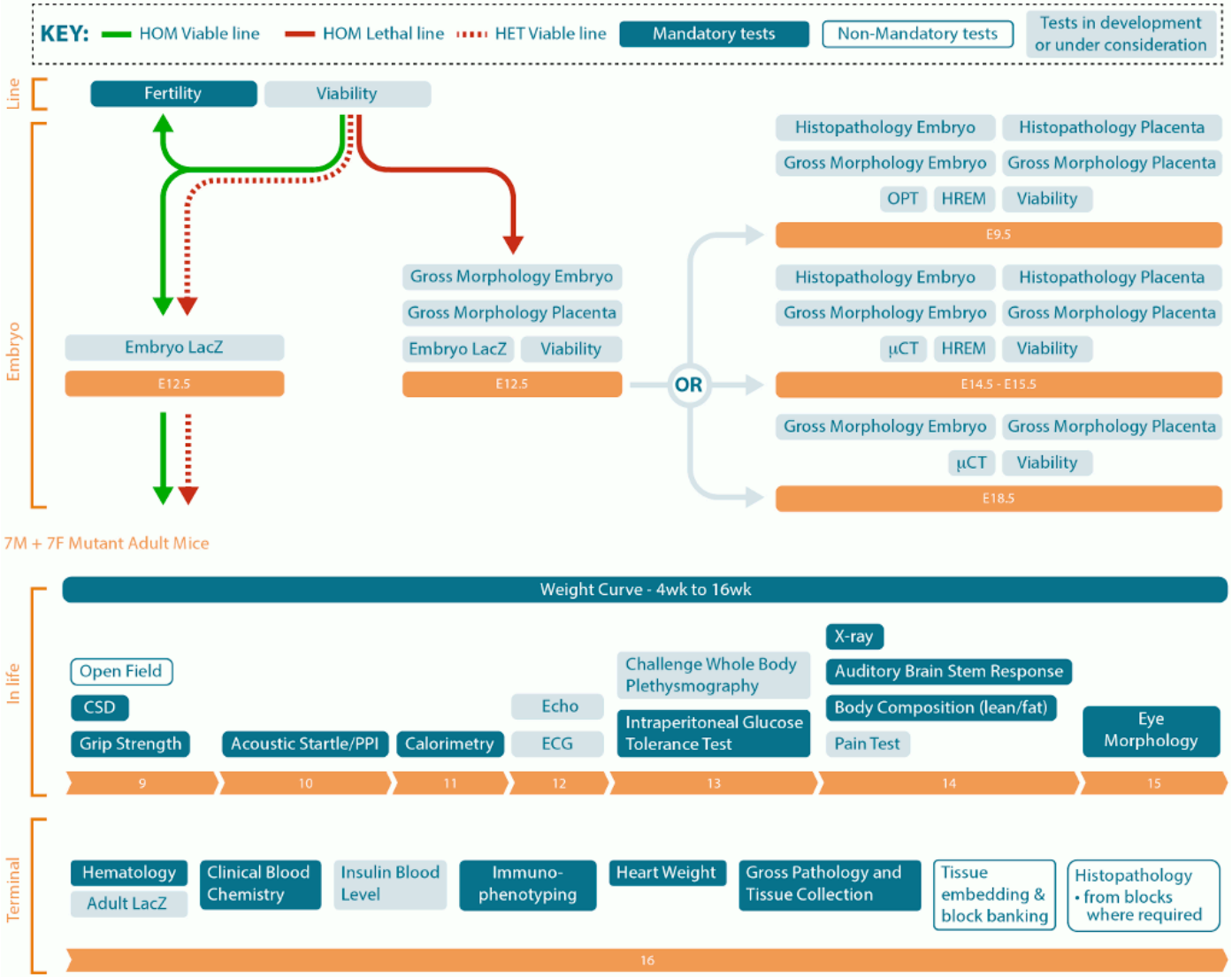
WC Skarnes *et al. Nature* **474**, 337-342 (2011) doi:10.1038/nature10163

IMPC

nature

MPI2

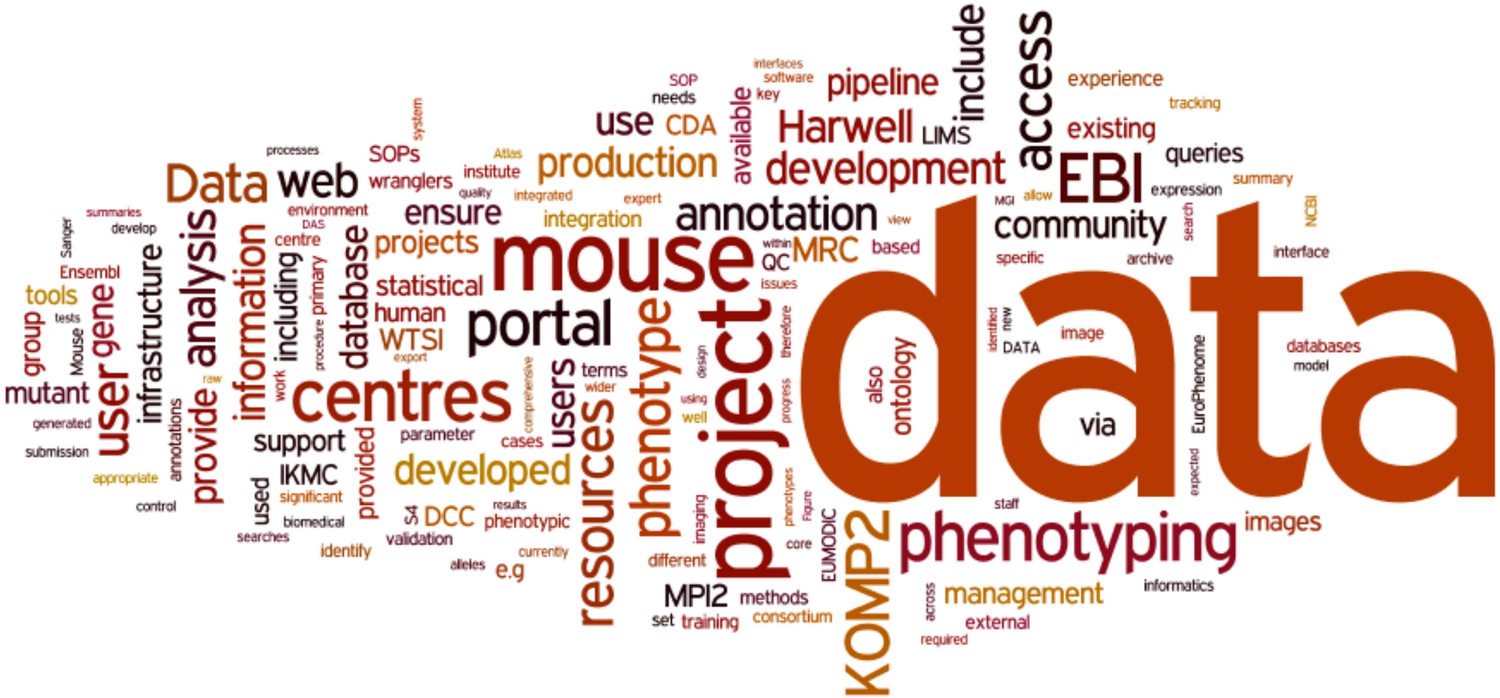
IMPC Core Phenotype Pipeline



IMPC

http://www.mousephenotype.org/impress/

MPI2


data
mouse
project
phenotyping
centres
portal
annotation
development
access
EBI
Harwell
pipeline
include
community
phenotype
resources
KOMP2
Data
web
analysis
information
user
gene
infrastructure
production
developed
use
CDA
ensure
projects
statistical
human
WTSI
database
including
provide
support
IKMC
DCC
MRC
LIMS
existing
experience
queries
expression
available
integration
tools
group
mutant
Ensembl
users
management
images
MPI2
methods
training
consortium
external
via
ontology
also
EuroPhenome
databases
SOPs
wranglers
institute
terms
parameter
cases
validation
identify
phenotypic
different
set
informatics

IMPC

MPI2

High Throughput Phenotyping Workflow

Mouse clinics

- Phenotype mouse following SOPs
- Generate data
- annotations

Pheno-DCC

- Import data
- QC data
- Export data to CDA

EBI Data warehouse

- Store data
- Analyse
- Integrate
- Disseminate

IMPC portal

- Search
- Browse
- Download results
- Link back to original data

wellcome trust sanger institute

MRC HARWELL

EMBL-EBI

IMPC

MPI2

Statistical Analysis Overview

- Goal: Associate Genotype to Phenotype via stats
- Statistical analysis depends on the experimental workflow
- Built statistical analysis platform on expected workflows
  - Fisher Exact Test
  - Linear regression- Mixed Model
- Improving analysis is active area of research
- Adjust to workflow at IMPC centers
- Versioning of statistical analysis
- Involvement of IMPC members

IMPC

MPI2

High Throughput Phenotyping Workflow

IMPC data coordination center

Mouse strain tracking (Sanger)

Data import, QC, prelim analysis (Harwell)

Phenotype associations (EMBL-EBI)

Disseminate Information
Website / Raw data

EBI statistical pipeline

Load data

Index experimental results

Calculate statistics

Create MP associations

IMPC

MPI2

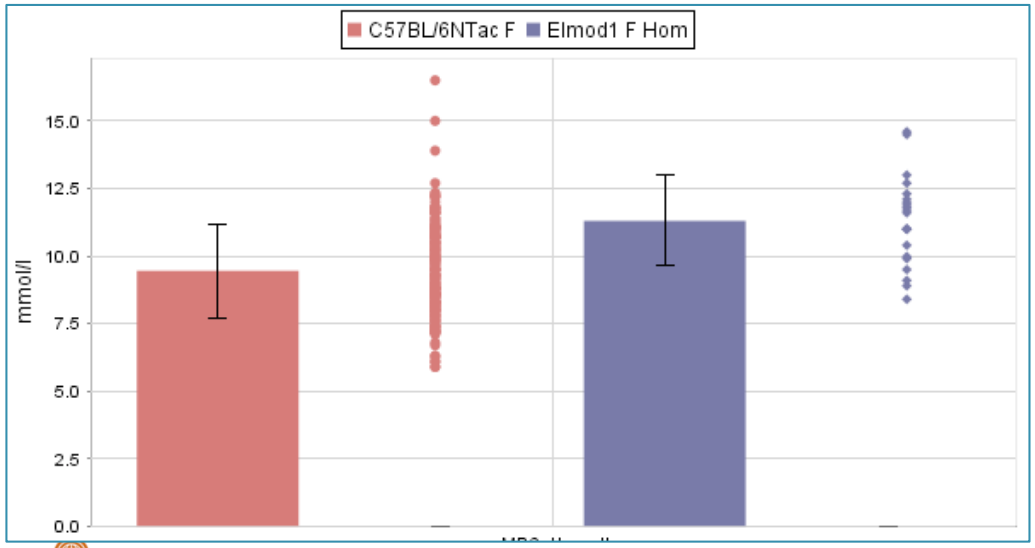
Procedure to Phenotypes

Clinical Blood Chemistry Procedure

	Version	Type	Req. Upload	Req. Analysis	Annotation	Option	Unit	Data Type
Sodium	1	simpleParameter	x	x	x		mmol/L	FLOAT
Potassium	1	simpleParameter	x	x	x		mmol/L	FLOAT
Chloride	1	simpleParameter	x	x	x		mmol/L	FLOAT
Urea	1	simpleParameter	x	x	x		mmol/L	FLOAT

Blood Urea Parameter

Parameter	Observation	Relative Value	MP Term	MP ID
Urea	mMol concentrations	ABNORMAL	abnormal blood urea nitrogen level	MP:0005265
Urea	mMol concentrations	INCREASED	increased blood urea nitrogen level	MP:0005565
Urea	mMol concentrations	DECREASED	decreased blood urea nitrogen level	MP:0005566

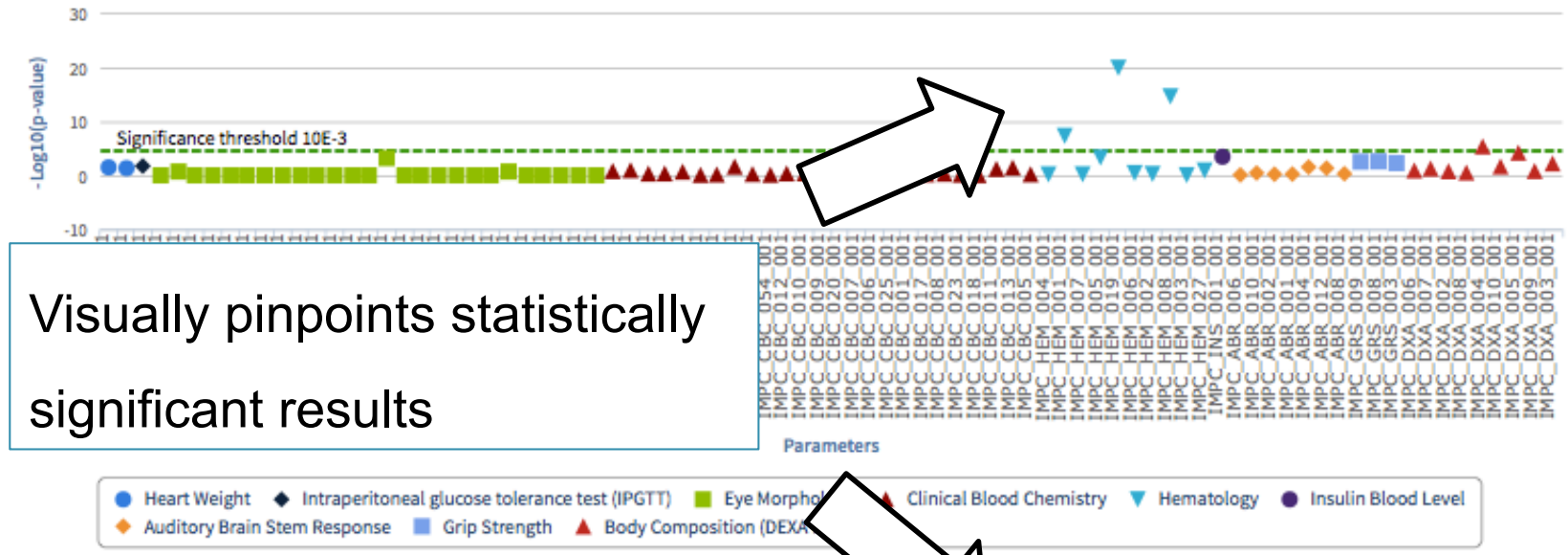


MP:0005565
increased blood urea nitrogen level

IMPC

MPI2

Statistical Analysis Overview



Visually pinpoints statistically significant results

Total number of results: 94

Procedure	Parameter	Data type	Zygosity	P-value	Status	Graph
Hematology	Mean platelet volume	unidimensional	homozygote	0.0	Success	
Hematology	Platelet count	unidimensional	homozygote	2.10942E-15	Success	
Hematology	White blood cell count	unidimensional	homozygote	4.71196E-8	Success	
Body Composition (DEXA lean/fat)	Bone Mineral Density (excluding skull)	unidimensional	homozygote	4.17771E-6	Success	
Body Composition (DEXA lean/fat)	Bone Mineral Content (excluding skull)	unidimensional	homozygote	6.11604E-5	Success	
Insulin Blood Level	Insulin	unidimensional	homozygote	3.48163E-4	Success	

Data Available on the IMPC Portal (June 2014)

- Preliminary results from statistical analysis available for **846** submitted knockout strains (QC in progress)
- **293** knockout strains QC'ed with complete data:
  - > **12M** data points measuring parameters from standardised operating procedures defined in IMPReSS
  - Normal fertility and viability
- > **98,000** images from the Wellcome Trust Mouse Genetics Program
- Number of line submitted will grow over time

IMPC

MPI2

IMPC and Imaging

- Understanding of in vivo gene function
- Identify novel genes
- Provide models of human disease
- Identification and highlight of unique anatomical structures expressing a targeted gene
- Detection and quantification of morphological and histological phenotypes (e.g. skeletal abnormalities)
- 30% of mutant strains are embryonic lethal, subviable
- Multiple modalities: HREM, OPT, micro-CT
- Images are linked to MP and MA ontology terms

IMPC

MPI2

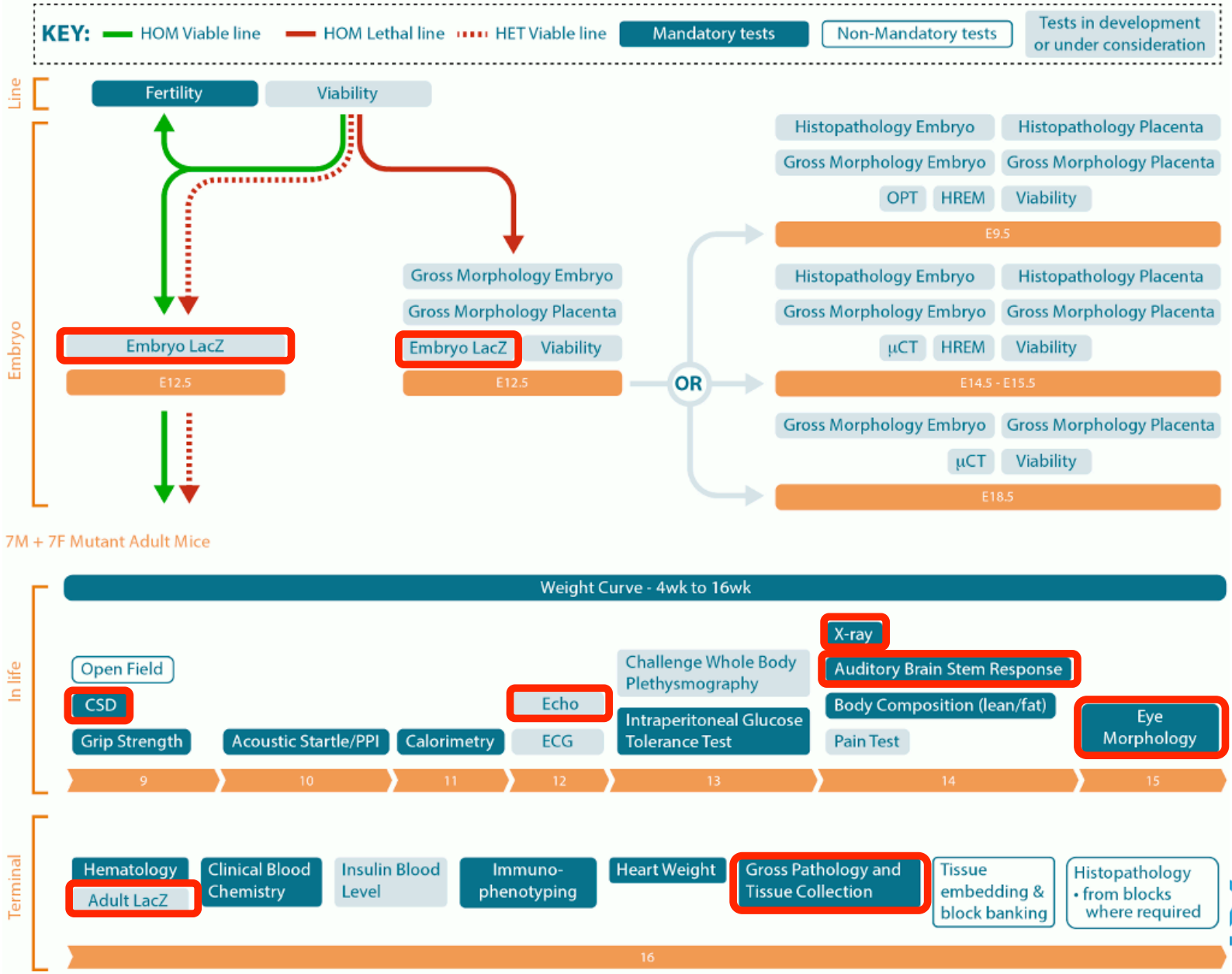
Use Cases

- Highlight morphological phenotypes in adult and embryo
- Visualise gene expression patterns (LacZ)
- Adult procedures (X-Ray, eye morphology, LacZ)
- Embryonic lethal mice: 3D imaging (HREM, OPT, micro-CT)
- Show images related to KO genes
- Retrieve images related to publication
- Compare mutants to baseline (wild-type images, Atlases)

IMPC

MPI2

Current pipeline schematic: 2D images



Developing Use Cases for 2D images

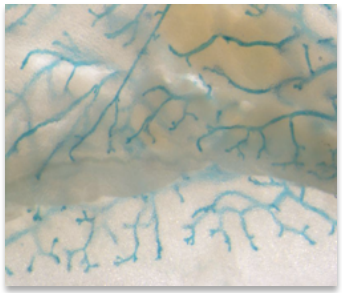
- **Wellcome Trust Sanger Institute Mouse Genetics Program**
- Rich resource of high-throughput phenotype images
- Test imaging import infrastructure
- Test representation of the information on the web
- **> 98,000** images to date (June 2014)
- Mouse anatomy and Mammalian phenotype ontology annotations
- Control mouse images available

IMPC

MPI2

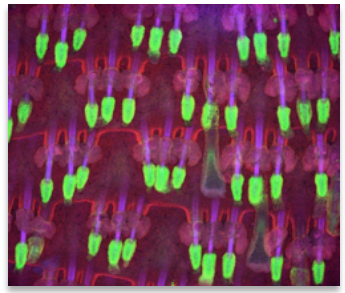
Mouse Genetics Program Imaging

High-Throughput
X-ray phenotyping

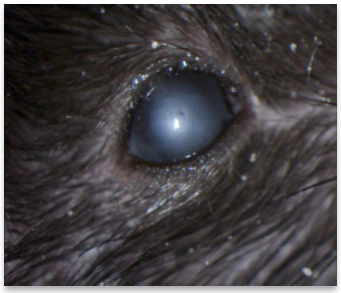


Wholemount
expression

Histology slide

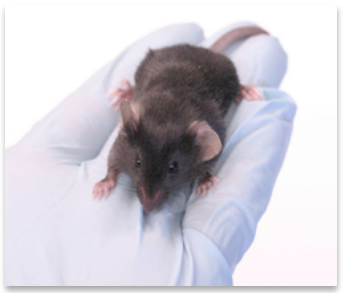


Tail epidermis
wholemount



Eye morphology

Embryo
morphology

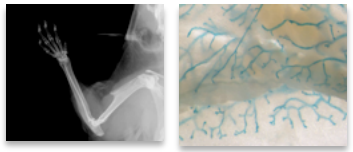


Dismorphology

IMPC

MPI2

Use case: Wellcome Trust Sanger Institute



Sanger MGP LIMS/FTP site

- Allele/strain information
- Procedure
- Annotations
- Images

KOMP2 database

- Import and store images, genotypes, procedures
- Index annotations for fast retrieval

Dissemination

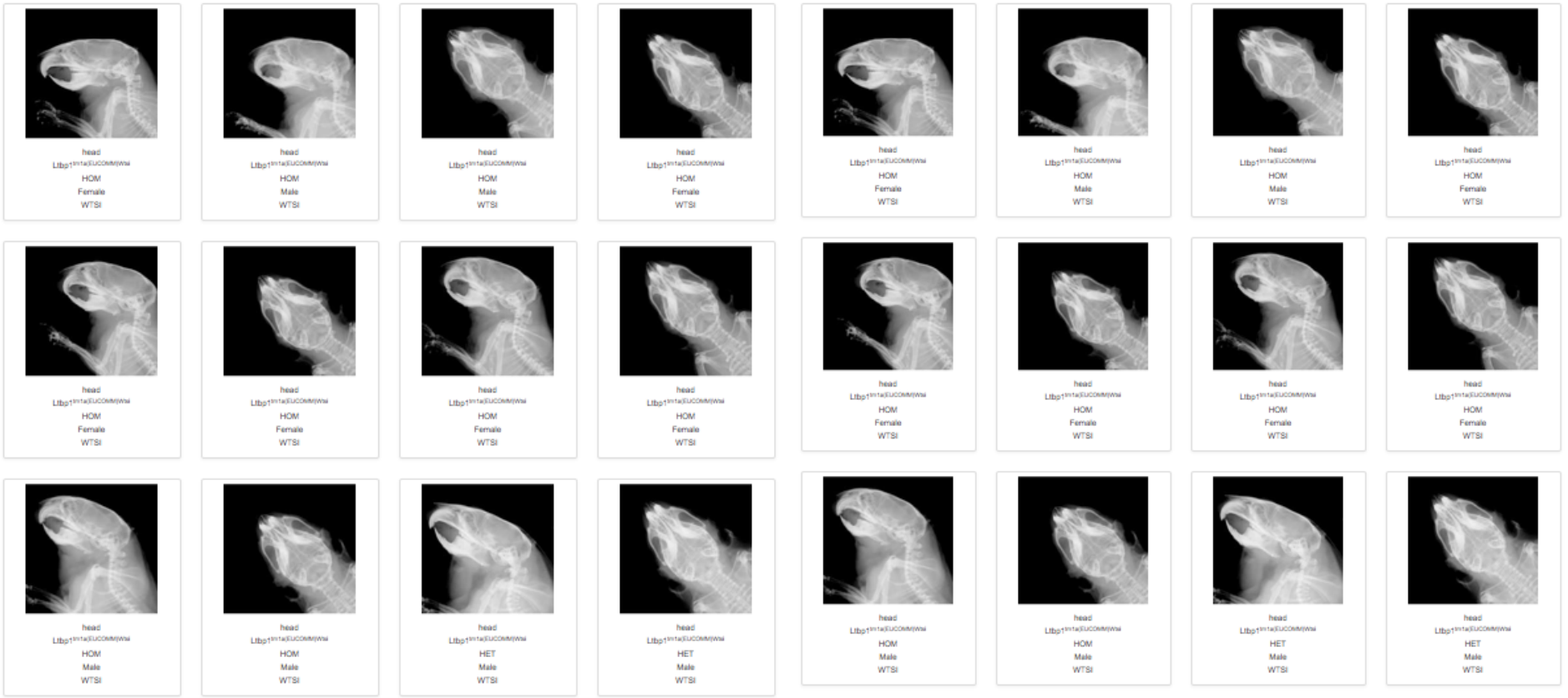
- Present images with annotations on portal
- Visible on gene, anatomy, phenotype page

IMPC

MPI2

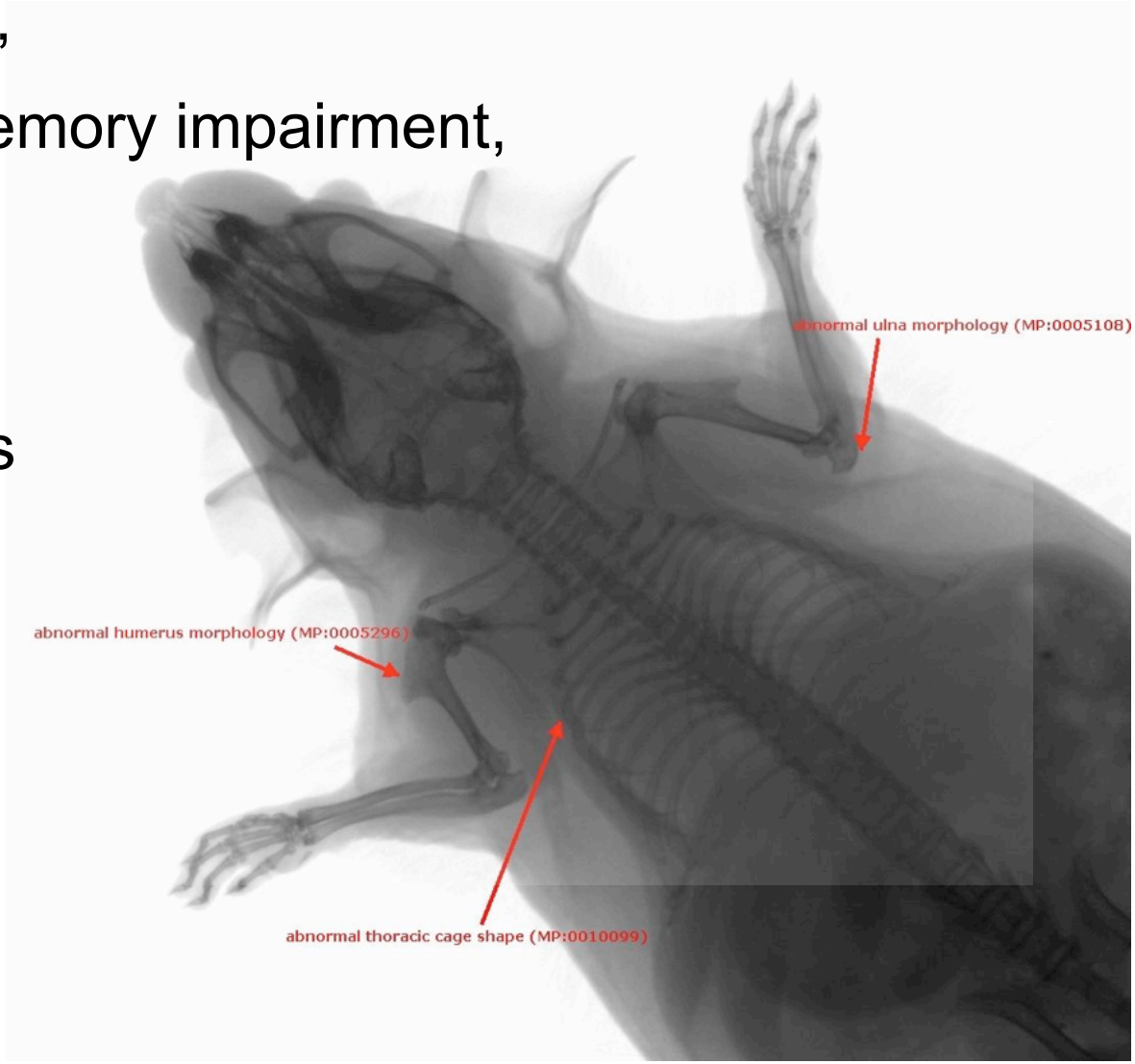
High-Throughput X-ray Phenotyping

- Detecting Skeletal abnormalities (DEXA scan)
- E.g. Ltbp1, Camurati-Engelmann disease (CED)



Cenpj$^{-/-}$ Phenotypes – Seckel Syndrome

- intrauterine dwarfism,
- microcephaly with memory impairment,
- ossification defects,
- ocular and
- skeletal abnormalities



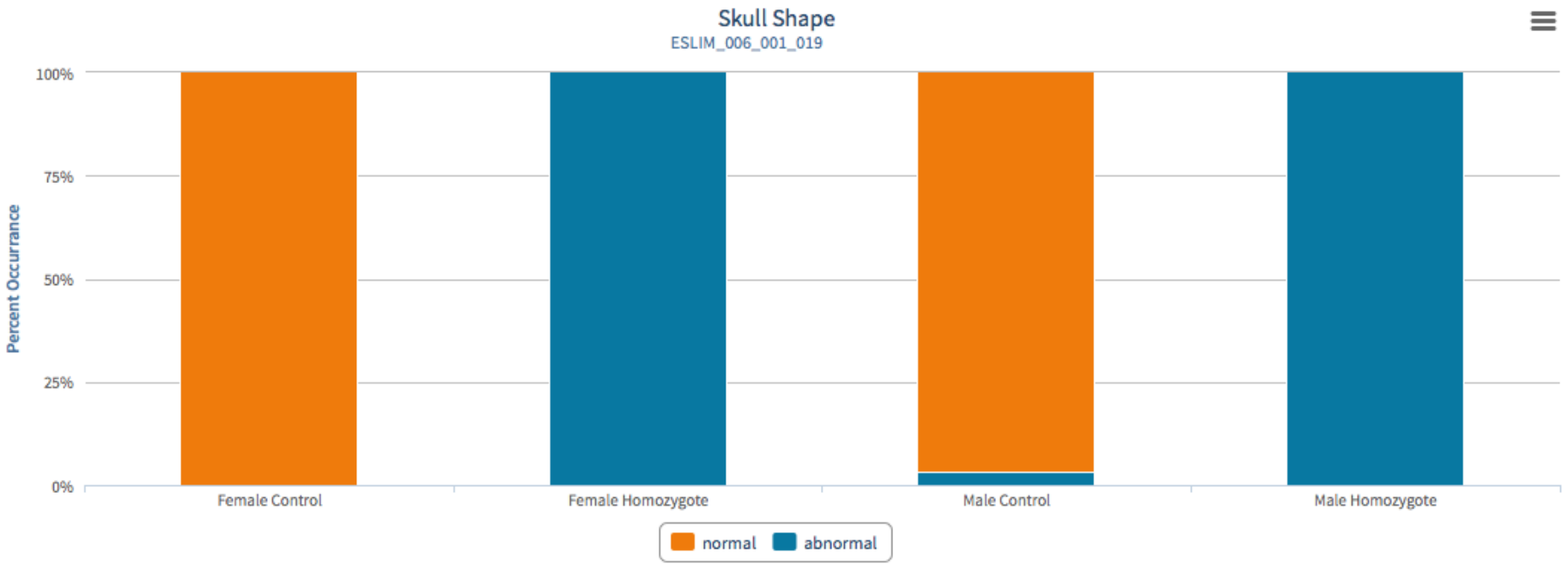
Disruption of mouse Cenpj, a regulator of centriole biogenesis, phenocopies Seckel syndrome
Rebecca McEntyre et al. PLoS Genet. 2012;8(11)

Statistics

- Compare KO animal results to baseline/WT
- Phenotypers report abnormality (using SOPs)
- Data are QC'ed and passed to EMBL-EBI for analysis and dissemination
- Example: Abnormal Cranium Morphology
- Fisher-Exact Test (here 2x2 contingency table)
- http://beta.mousephenotype.org/data/genes/MGI:2444584

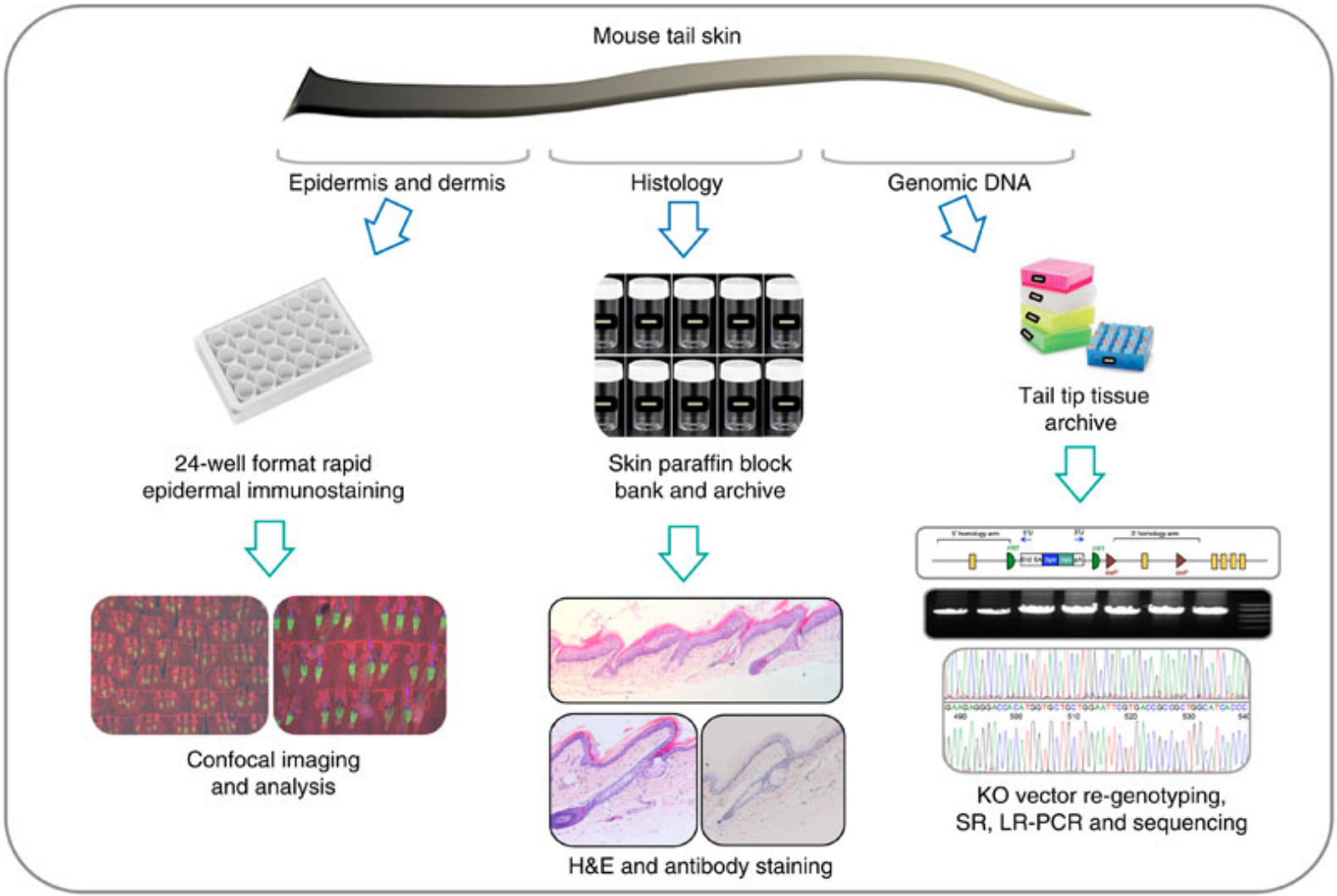
Allele - Mysm1$^{tm1a(KOMP)Wtsi}$

Background - involves: C57BL/6 Phenotyping Center - WTSI Pipeline - EUMODIC Pipeline 1



Control/Hom/Het	normal	abnormal	P Value	Max Effect
Female Control	120	0		
Female Homozygote	0	5	4.26382E-9	1.0
Male Control	115	4		
Male Homozygote	0	8	3.69308E-10	0.97

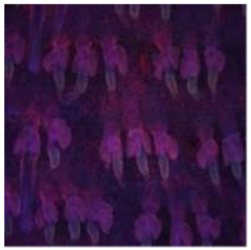
Krt76$^{-/-}$ Skin Phenotypes



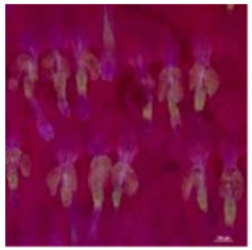
Novel skin phenotypes revealed by a genome-wide mouse reverse genetic screen
Nature Communications 5, Article number: 3540
doi:10.1038/ncomms4540

Krt76$^{-/-}$ Skin Phenotypes

- Flaky tail skin
- *lacZ* reporter expression in footpad
- Confocal epidermal wholemount images of Krt76-/-



Krt76$^{tm1a(KOMP)Wtsi}$ tail skin, abnormal epidermis stratum basale morphology, abnormal hair cycle, abnormal sebaceous gland morphology, abnormal epidermis stratum basale morphology, abnormal hair cycle, abnormal sebaceous gland morphology, HOM, Male, WTSI



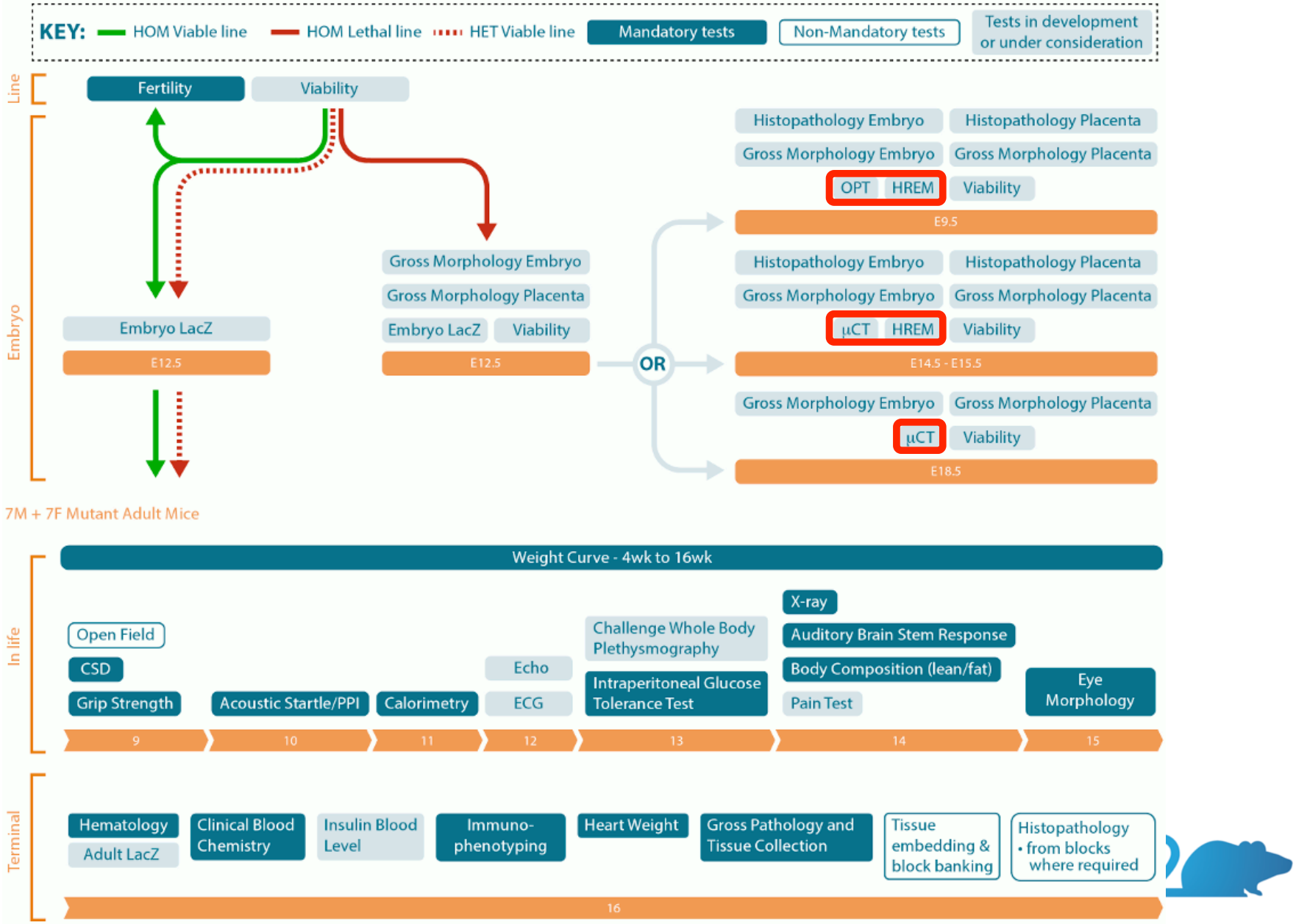
Krt76$^{tm1a(KOMP)Wtsi}$ tail skin, abnormal hair cycle, abnormal sebaceous gland morphology, abnormal epidermis stratum basale morphology, abnormal hair cycle, abnormal sebaceous gland morphology, abnormal epidermis stratum basale morphology, HOM, Male, WTSI

Krt76$^{tm1a(KOMP)Wtsi}$ tail skin, abnormal sebaceous gland morphology, abnormal epidermis stratum basale morphology, abnormal hair cycle, abnormal epidermis stratum basale morphology, abnormal hair cycle, abnormal sebaceous gland morphology, HOM, Male, WTSI

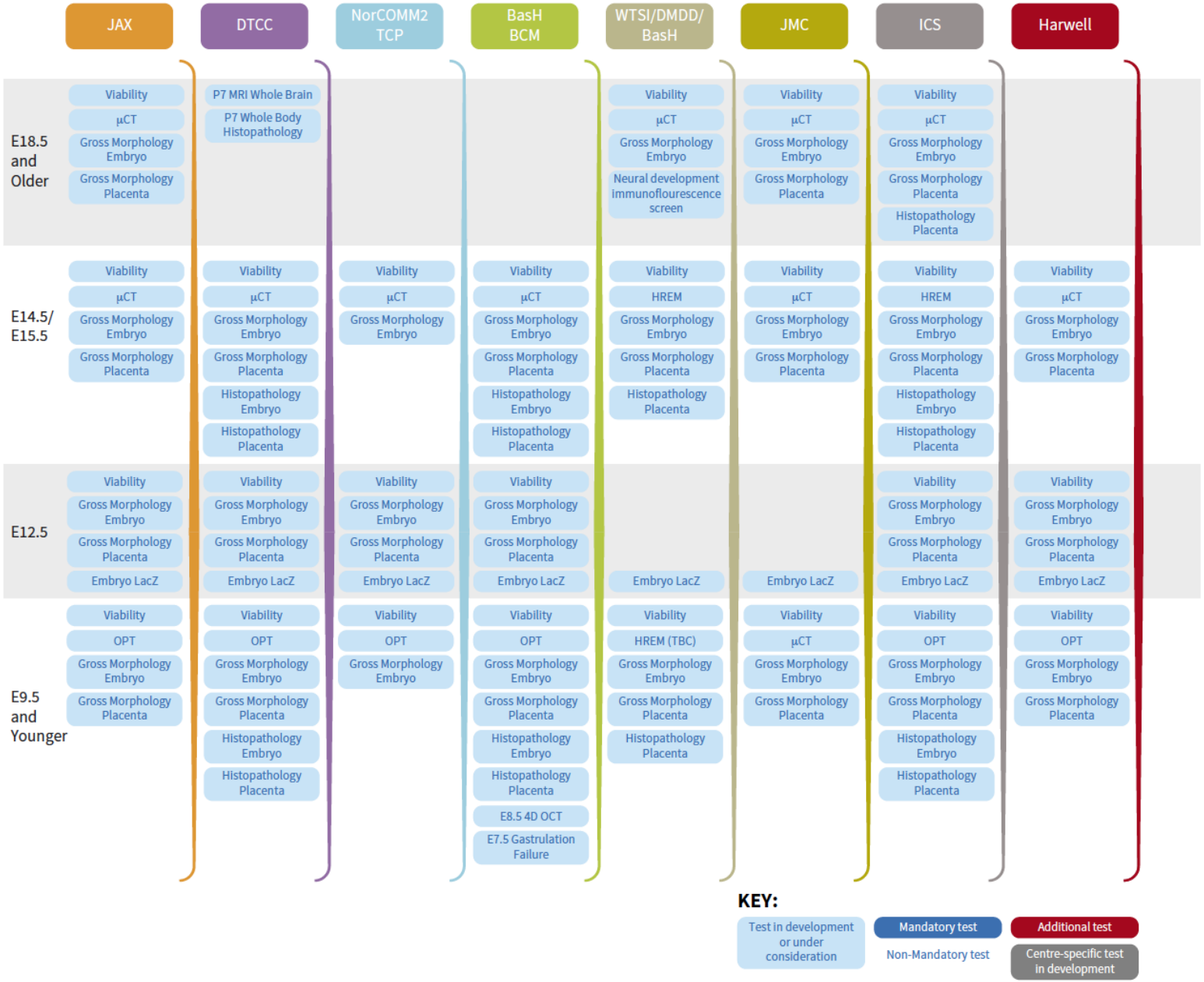
Krt76$^{tm1a(KOMP)Wtsi}$ tail skin, abnormal epidermis stratum basale morphology, abnormal hair cycle, abnormal sebaceous gland morphology, abnormal sebaceous gland morphology, abnormal epidermis stratum basale morphology, abnormal hair cycle, HOM, Male, WTSI

Novel skin phenotypes revealed by a genome-wide mouse reverse genetic screen
Nature Communications 5, Article number: 3540
doi:10.1038/ncomms4540

Current pipeline schematic: 3D images



Tests per center/consortium

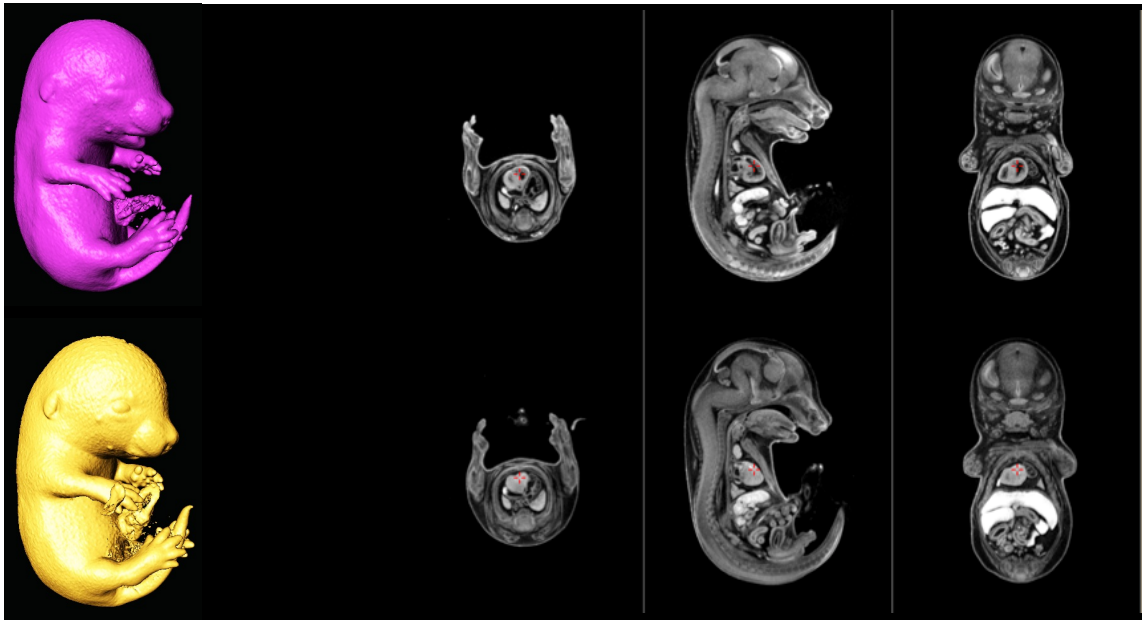


Three-dimensional (3D) Volumetric Imaging

- Whole embryo coverage
- Single gene mutations can cause multi-organ abnormalities that single tissue or two-dimensional (2D) sections can easily miss
- Resulting image comprises digital data that can be manipulated for optimal viewing

WT

KO



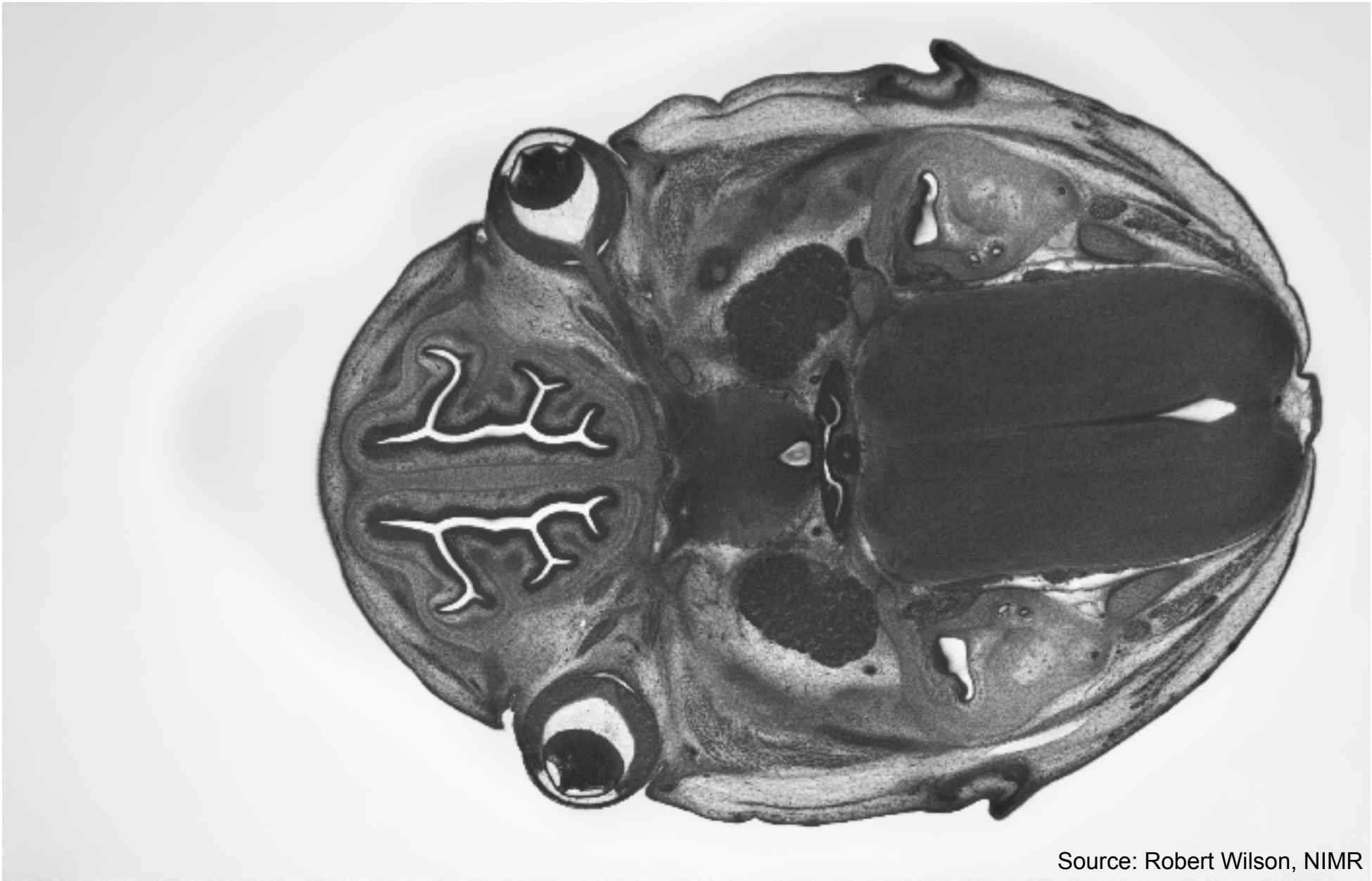
Source: Michael Wong, TCP

DMDD

Deciphering Mechanisms of Developmental Disorders

- Goals:
  - Advance understanding of embryo development
  - Identify models of developmental disease
  - Identify novel disease genes
- Systematic phenotyping of embryonic lethals (WTSI lines)
  - Imaging: µCT, HREM
  - **Manual** annotations
  - Dissemination: embryoimaging.org

Tim Mohun et al.
Dis Model Mech. 2013 May;6(3):562-6



Source: Robert Wilson, NIMR

IMPC

MPI2

HREM Embryo data sets

E14.5 wt

Wild-type: 93 embryos

7 stages

NIMR:Parkes and C57BL/6 genetic backgrounds

297,090 images

Psat1 -/-

Mutant lines: 91

- (383 embryos; 252 homozygous for a mutant allele)

E14.5 stage

7 genetic backgrounds

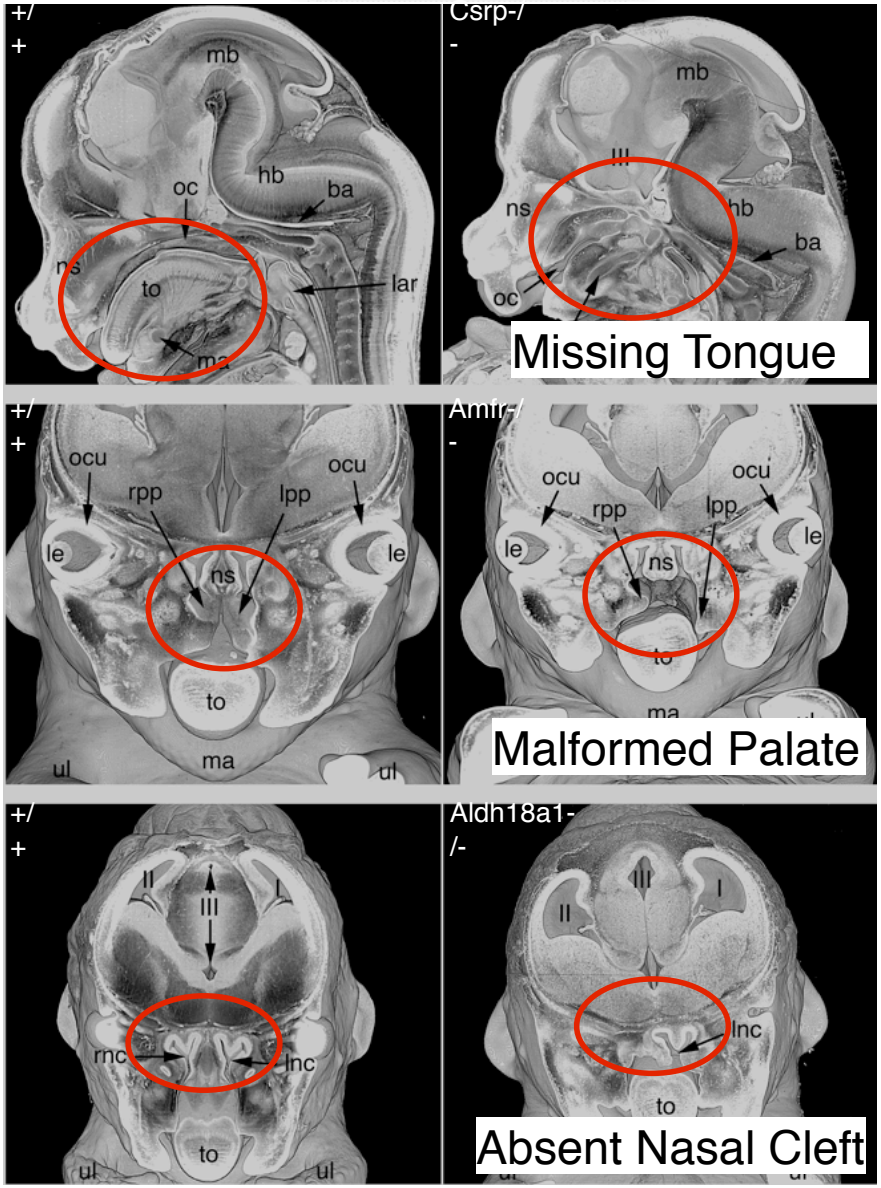
1,203,808 images

Tim Mohun et al.
Dis Model Mech. 2013 May;6(3):562-6

Pilot HREM embryo screen

30 lines (122 embryos)
120 different defects

gross organ malformation → subtle vascular/nerve defect



Source: Robert Wilson, NIMR

Micro-CT Embryo data sets

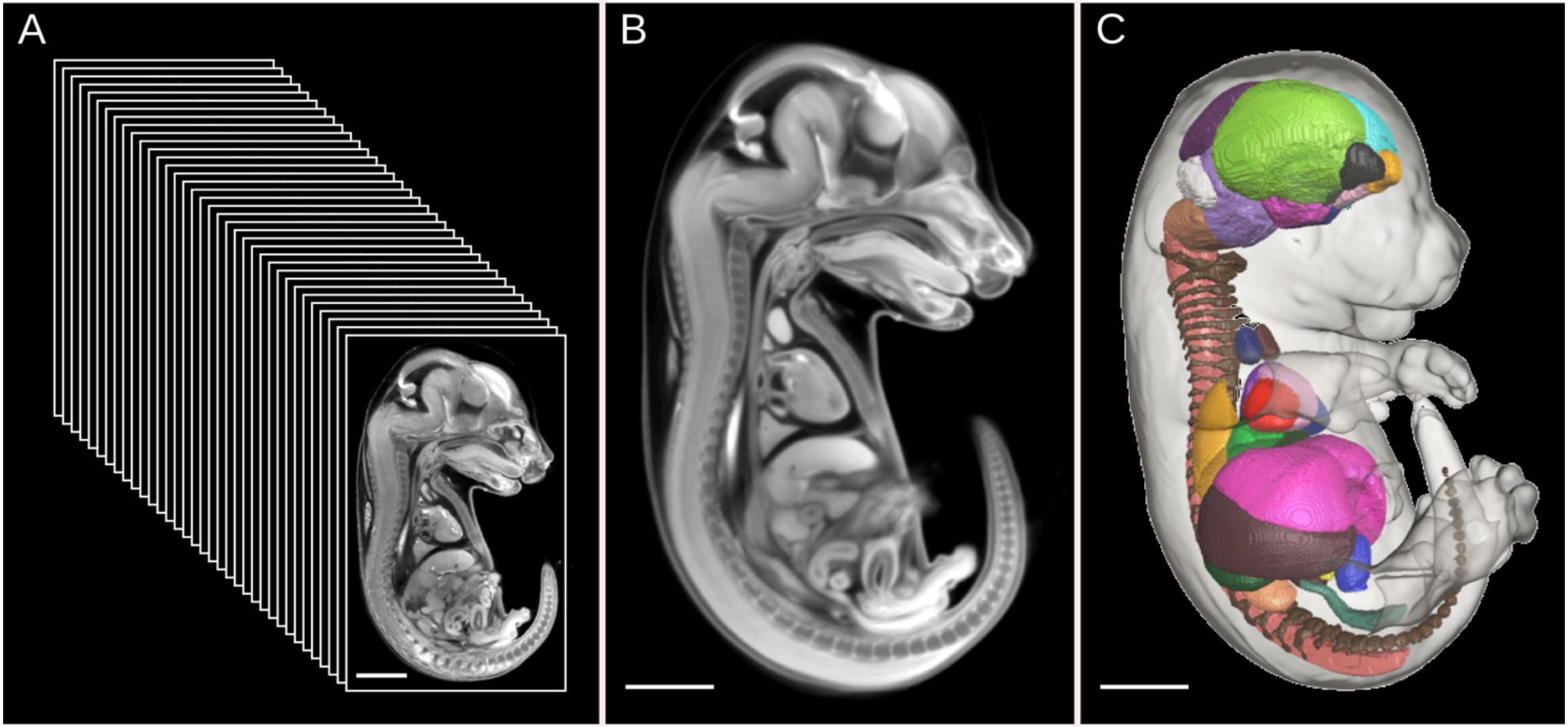
- Mark Henkelman's group - TCP (University of Toronto)
- Automatically quantify morphological information of mouse embryos from 3D imaging
- Characterize the variability within the normal population
- 15.5 dpc C57BL/6J mouse embryo
- A reference 3D atlas has been developed of 48 structural volumes
- http://www.mouseimaging.ca/technologies/mouse_embryo_atlas.html

Wong, M. D., Dorr, A. E., Walls, J. R., Lerch, J. P. and Henkelman, R. M. (2012). A novel 3D mouse embryo atlas based on micro-CT. Development 139, 3248–3256.

IMPC

MPI2

Building a 3D atlas of 48 structural volumes



Wong, M. D., Dorr, A. E., Walls, J. R., Lerch, J. P. and Henkelman, R. M. (2012).
A novel 3D mouse embryo atlas based on micro-CT. Development 139, 3248–3256.

IMPC

MPI2

Micro-CT Data Analysis

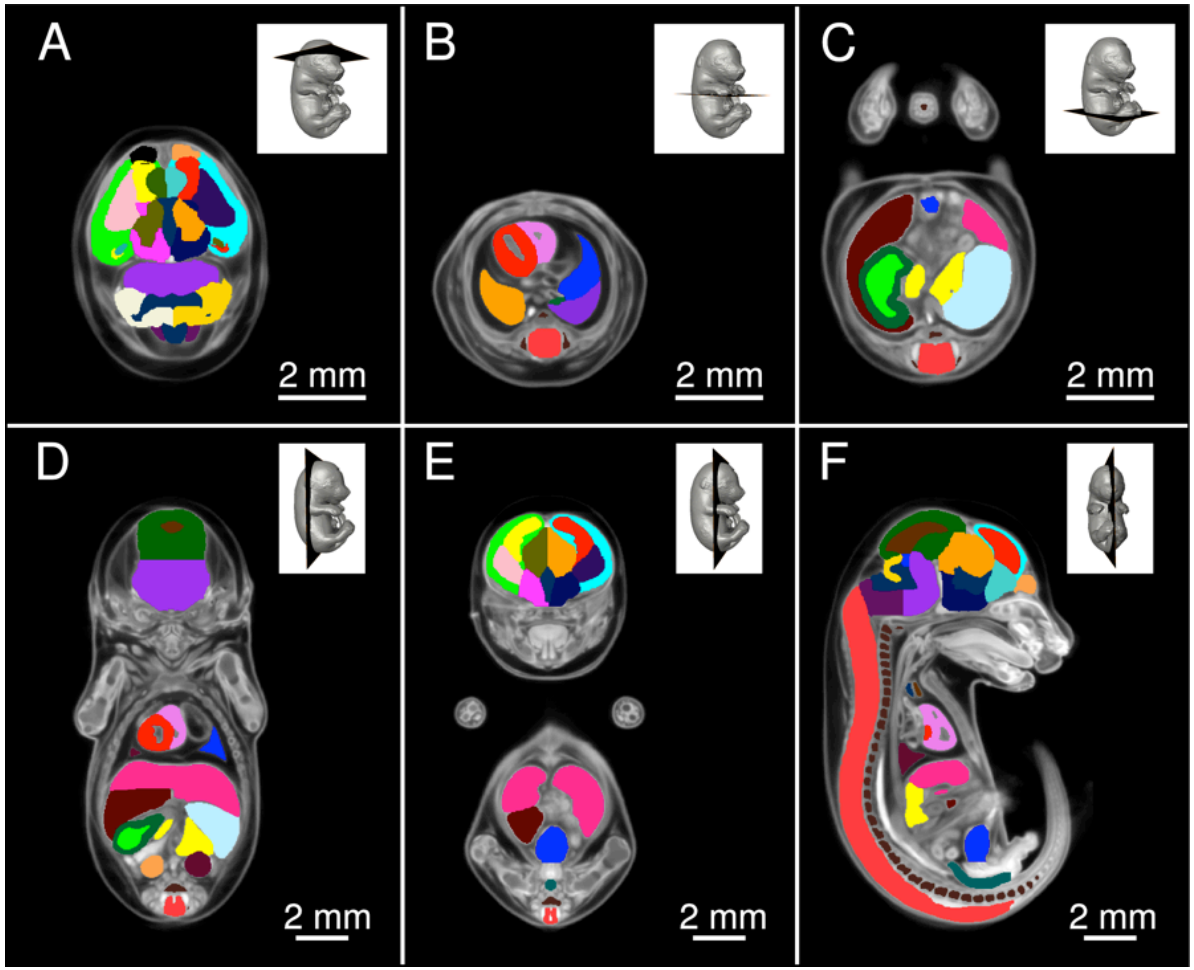
3 types analysis based on the results of the image registration:

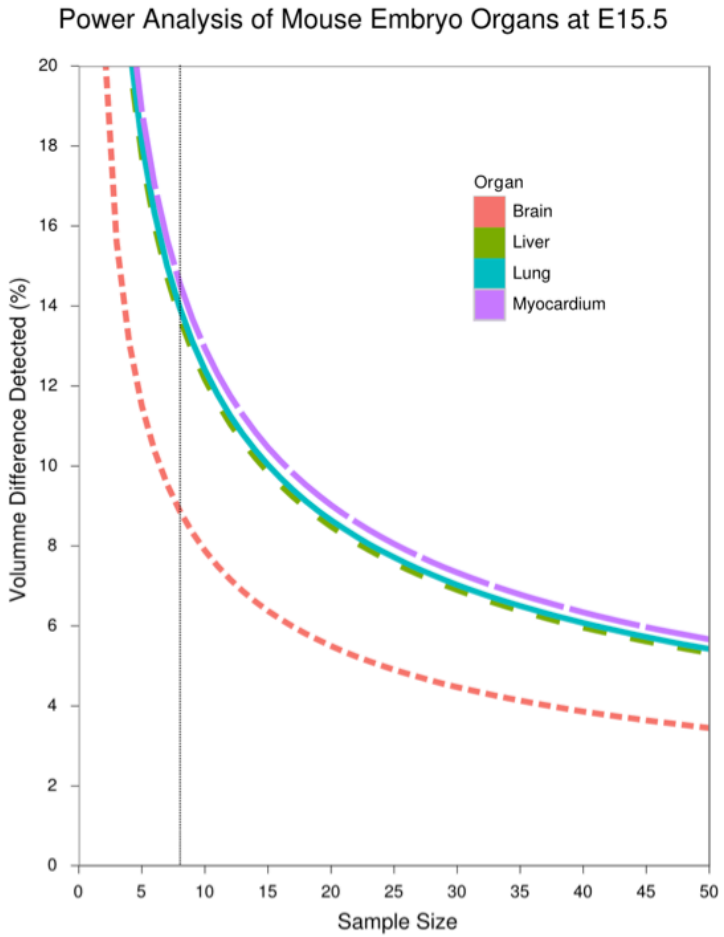
1) Missing structures, gross differences
   - Image intensity differences post registration
2) Local volume differences
   - Differences in the amount of deformation to reach the average image
3) Whole structure volume differences
   - Structure volumes from resampled segmented atlas

IMPC

MPI2

Volume differences of 9-14% can be detected by sample sizes of n=8



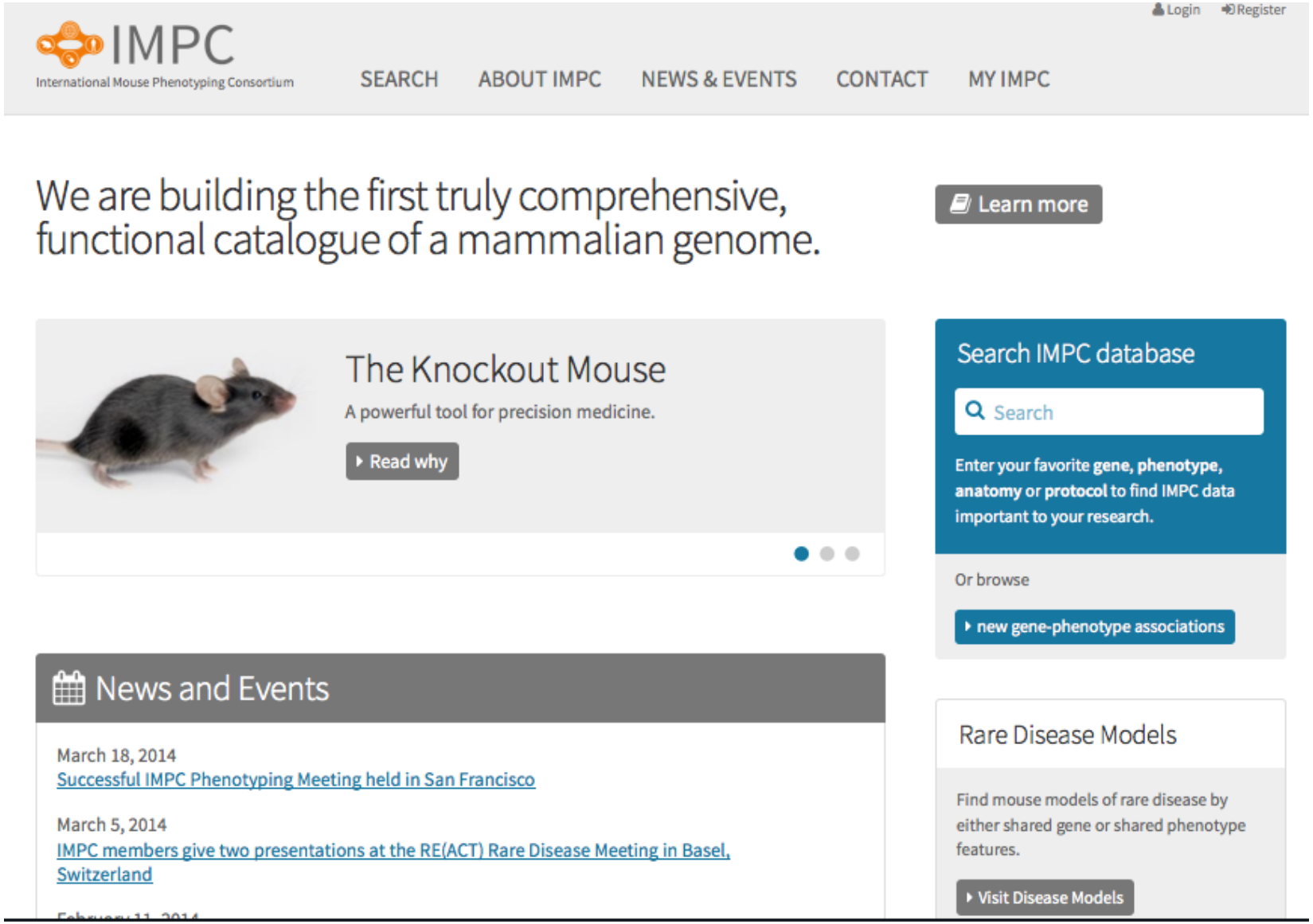


IMPC

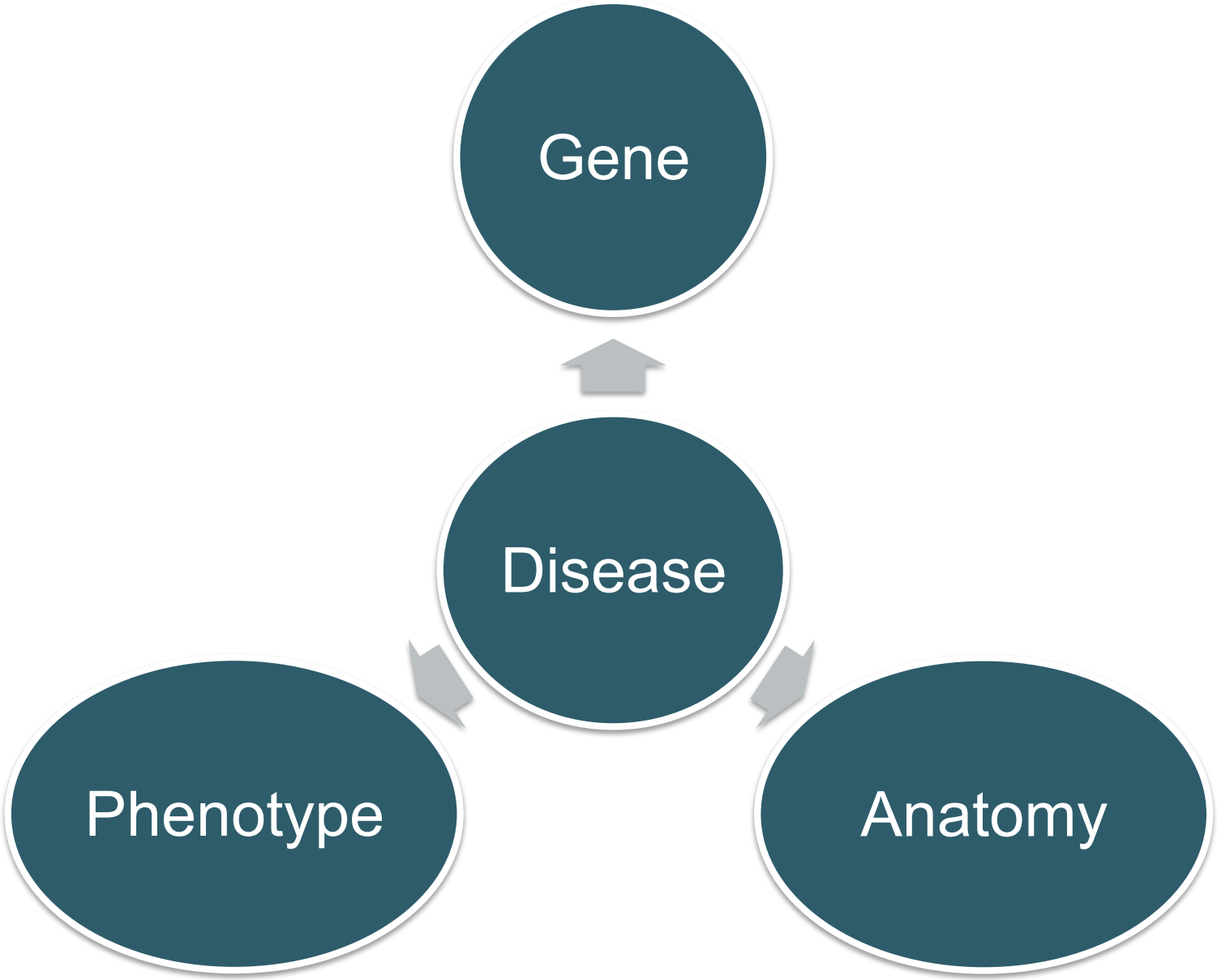
Wong, M. D., Dorr, A. E., Walls, J. R., Lerch, J. P. and Henkelman, R. M. (2012). A novel 3D mouse embryo atlas based on micro-CT. Development 139, 3248–3256.

MPI2

New Web Portal – beta.mousephenotype.org



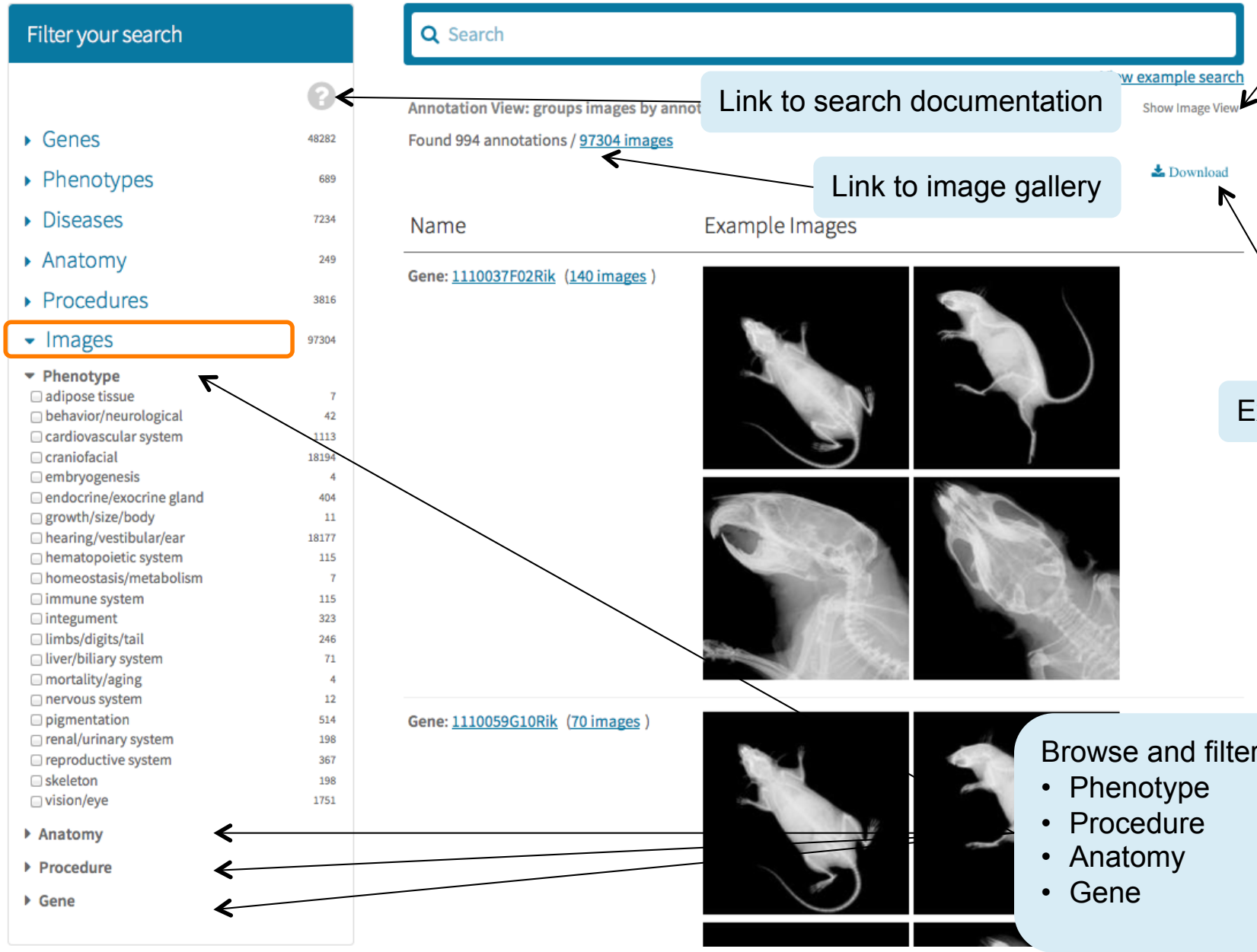
Portal Search beta.mousephenotype.org



IMPC

MPI2

Retrieving images on the portal



Grouping Images: by gene attributes (Cenpj)

Filter your search

- Genes 1
- Phenotypes 64
- Diseases 3
- Anatomy 0
- Procedures 0
- Images 230

Phenotype

- cardiovascular system 6
- craniofacial 74
- endocrine/exocrine gland 1
- hearing/vestibular/ear 74
- integument 1
- limbs/digits/tail 16
- pigmentation 1
- renal/urinary system 1
- reproductive system 6
- skeleton 7
- vision/eye 6

Anatomy

- extremity 113
- integumental system 1

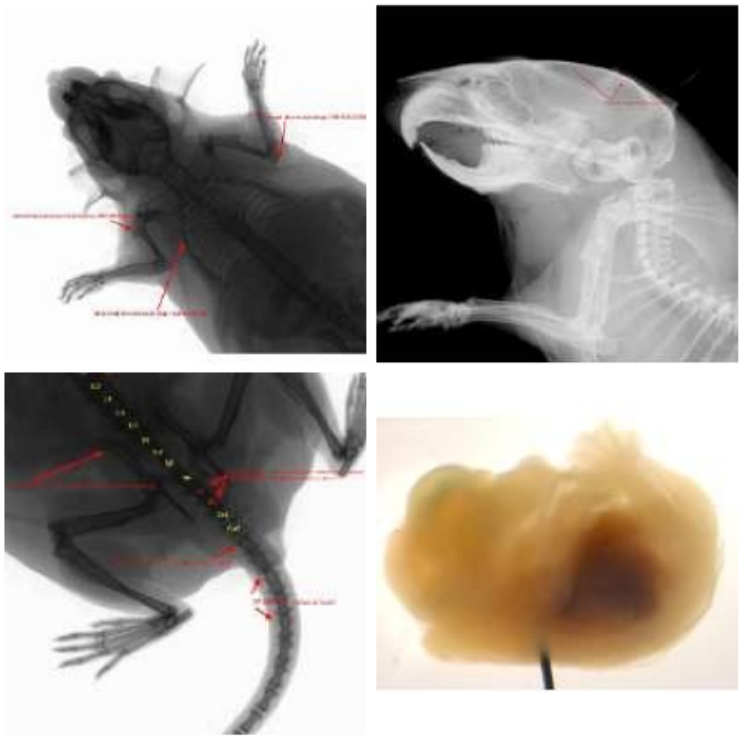
"cenpj"

View example search

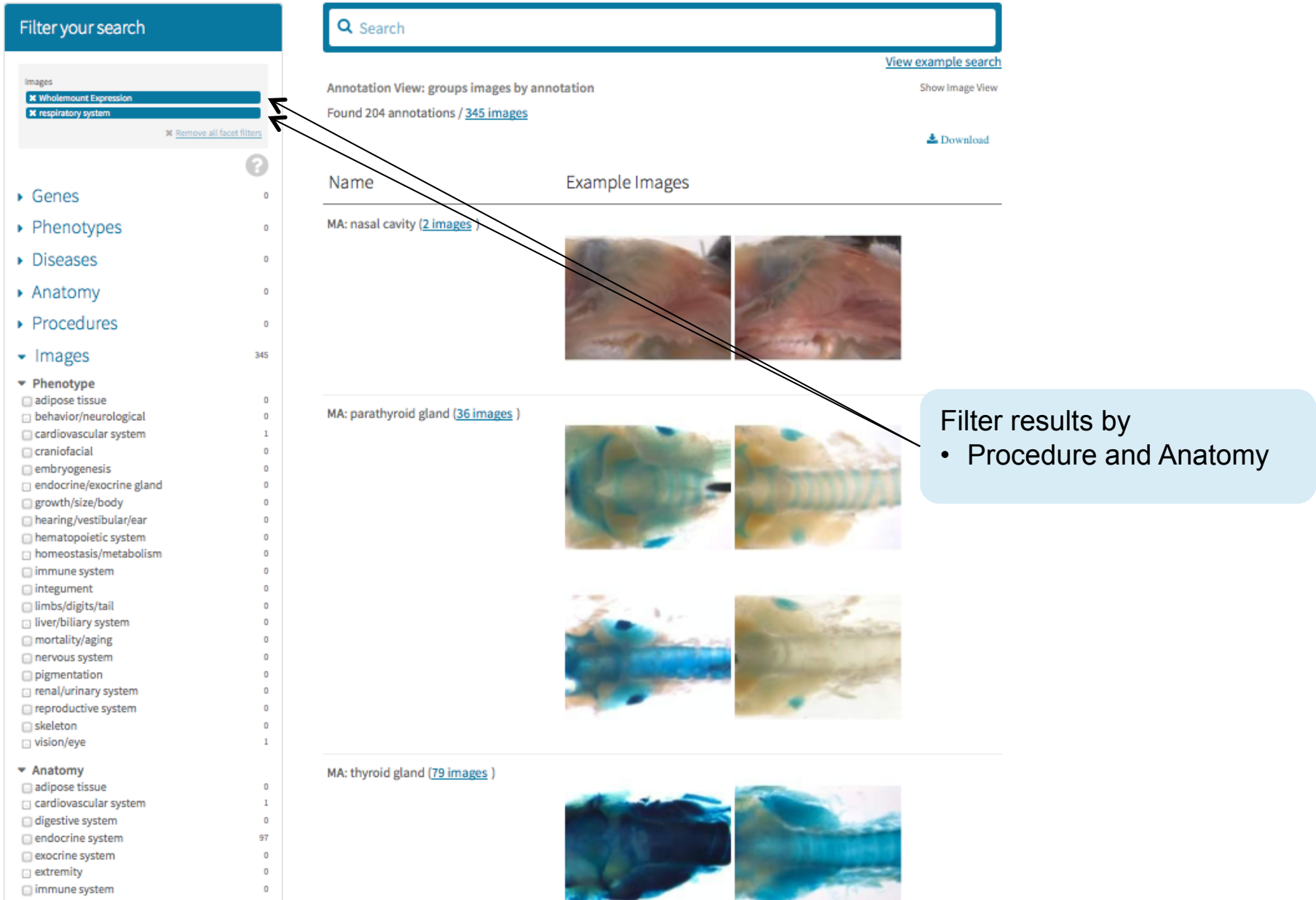
Annotation View: groups images by annotation Show Image View

Found 27 annotations / 230 images

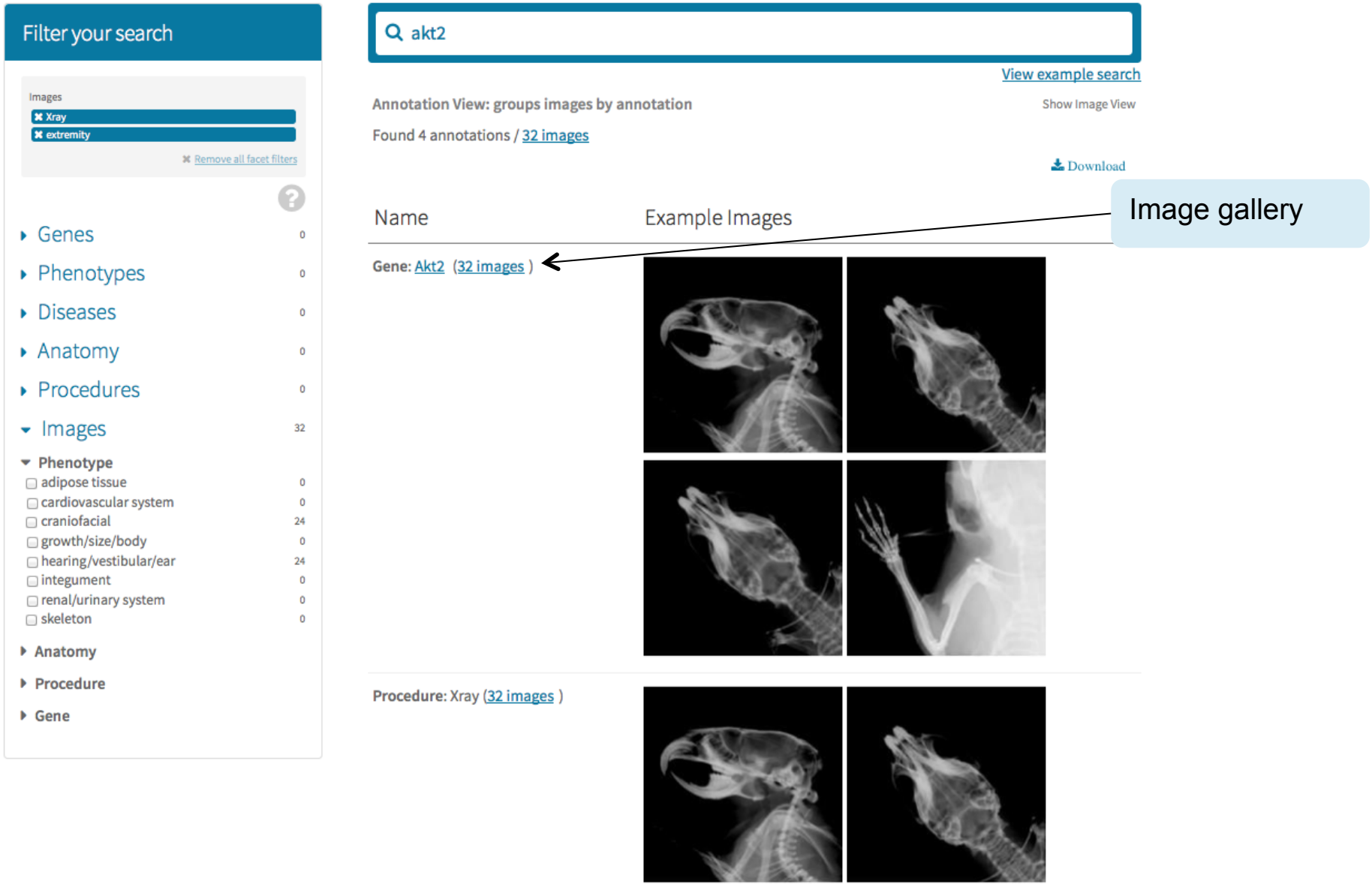
Download

Name	Example Images
Gene: Cenpj (229 images)	

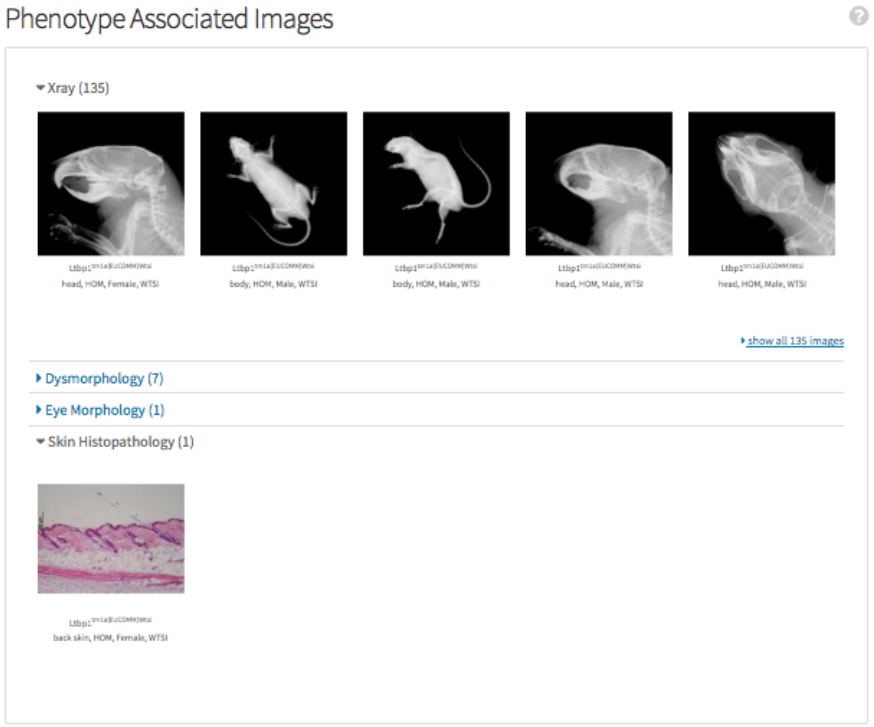
Filtering: “Wholemount expression” and “respiratory system”



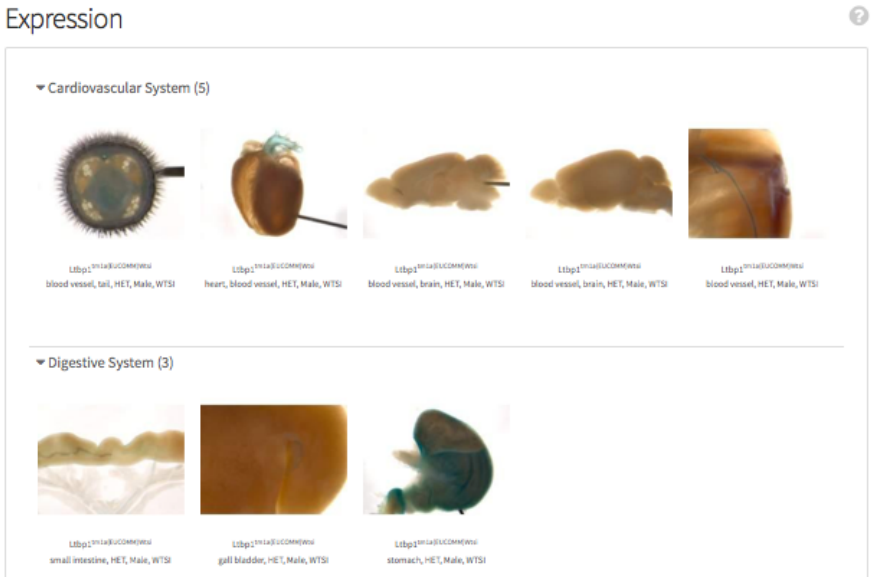
Filtering: “Akt2” and “X-ray” and “extremity”



Accessing images from the gene page



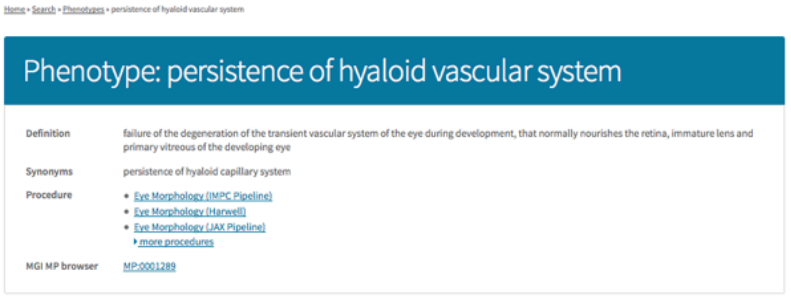
Phenotype images associated to the gene are presented test by test when available



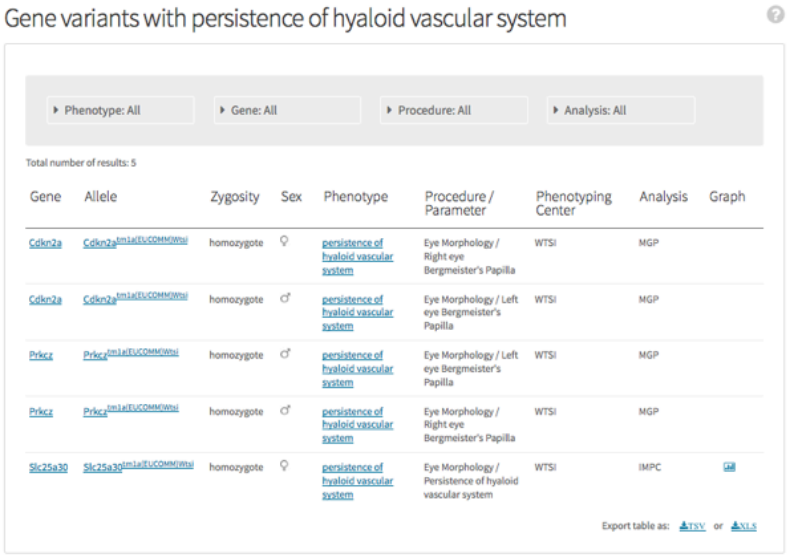
Wholemount expression images available from gene page

Abnormal Phenotype Page

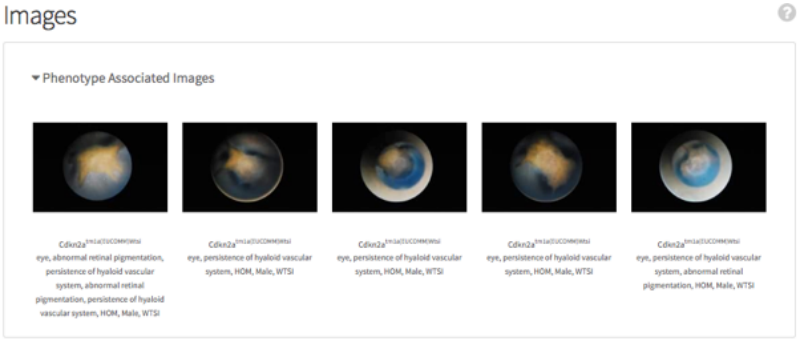
http://beta.mousephenotype.org/data/phenotypes/MP:0001289



Summary



Mutant mouse strains



Associated images

Challenges and Future Directions

- Linking back to original image provider:
  - Unique identifier for each image linking to specific region (e.g. NDPI viewer)
- Managing images at EMBL-EBI: OMERO 5 platform pilot study
- Providing tools to annotate the images after their transfer to EMBL-EBI
- Planning with GXD to exchange data (LacZ images)
- Display volumetric phenotypes at different embryonic development stage:
  - 3D plugin (Woolz IIP)
- Implement a reference Atlas (http://www.mouseimaging.ca/technologies/mouse_embryo_atlas.html)
- Exchange data with **PhenoImageShare**: federated resources to retrieve and annotate phenotype images from vertebrates and plants (BBSRC)

IMPC

MPI2

Acknowledgements- Funding: U54 HG006370

- MRC Harwell:
  - Ahmad Retha
  - Andrew Blake
  - Armida Di Fenza
  - Duncan Sneddon
  - Gagarine Yaikhom
  - Henrik Westerberg
  - Julian Atienza-Herrero
  - Luis Santos
  - Tanja Fiegel
  - Hilary Gates
  - Ann-Marie Mallon
  - Steve Brown
- TCP:
  - Michael Wong
- Welcome Trust Sanger Institute:
  - Alan Horne
  - Natasha Karp
  - Nida Jahangir
  - Richard Easty
  - Robert Wilson
  - Peter Matthews
  - David Melvin
  - Vivek Iyer
  - Bill Skarnes
  - Jacqui White,
  - Jeanne Estabel,
  - Anna-Karin Gerdin
- DMMD:
  - Robert Wilson
- European Bioinformatics Institute:
  - Chao-Kung Chen
  - Jeremy Mason
  - Jonathan W.G. Warren
  - Gautier Koscielny
  - Natalja Kurbatova
  - Ilinca Tudose
  - Mike Relac
  - Terry Meehan
  - Helen Parkinson
  - Paul Flicek
- NIH
  - Mark Moore
  - Colin Fletcher

IMPC

MPI2