IMPC

International Mouse Phenotyping Consortium

@impc

API Workshop - Disease section

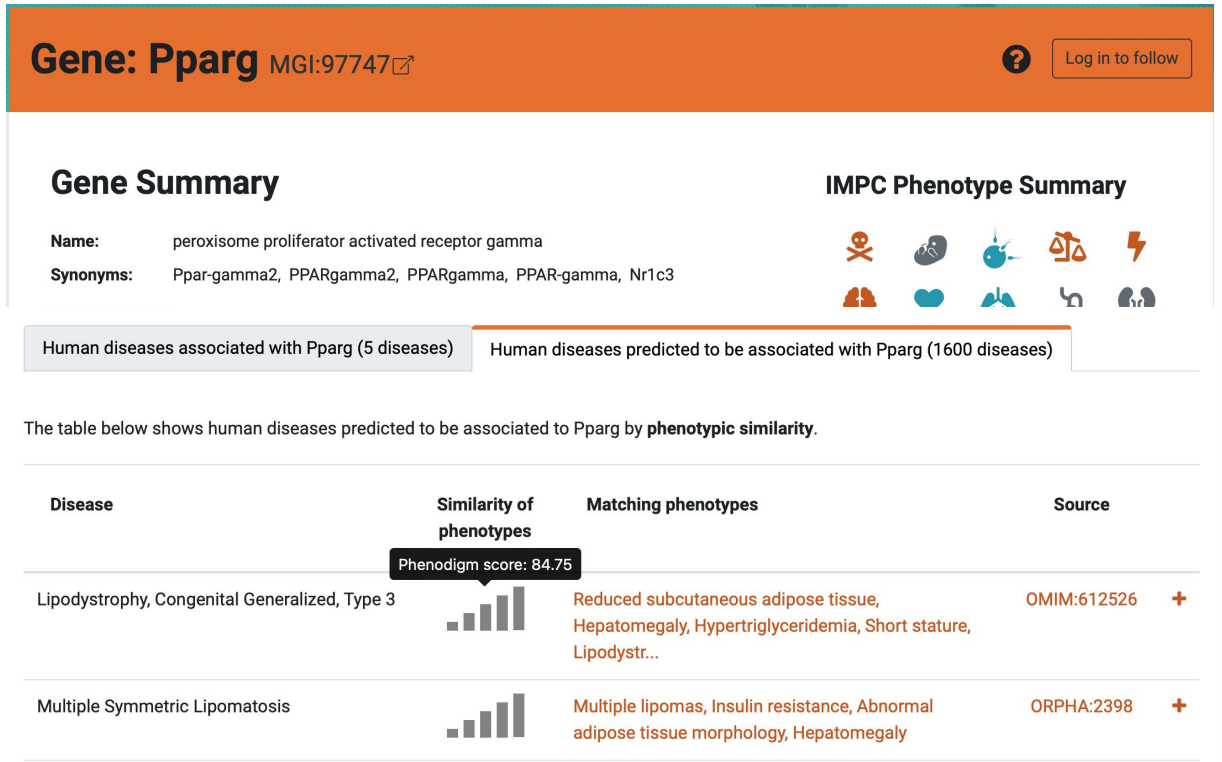
Phenodigm core

Diego Pava – Bioinformatician QMUL

mousephenotype.org

Human – mouse matching

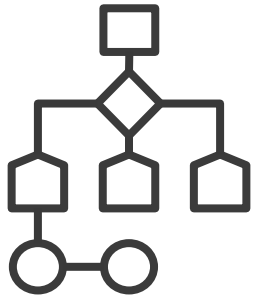
- Data analysis of the data generated by the IMPC pipeline
- Predict human diseases associated with a gene
- Aid in finding novel disease models
- PhenoDigm algorithm implementation



PhenoDigm and application within IMPC

- **P**henotype comparisons for **DI**sease **G**enes and **M**odels (PhenoDigm)
- This quantitative measure can tell us how well our mouse models recapitulate the clinical features of a disease [1]
- Ontologies: Mammalian Phenotype Ontology (MP) and Human Phenotype Ontology (HPO)
- Disease HPO terms aligned with MP terms and produce a **percentage similarity score**

PhenoDigm



JOURNAL ARTICLE

PhenoDigm: analyzing curated annotations to associate animal models with human diseases

Damian Smedley ✉, Anika Oellrich, Sebastian Köhler, Barbara Ruef, Sanger Mouse Genetics Project, Monte Westerfield, Peter Robinson, Suzanna Lewis, Christopher Mungall Author Notes

Database, Volume 2013, 2013, bat025, https://doi.org/10.1093/database/bat025
Published: 09 May 2013 **Article history** ▾

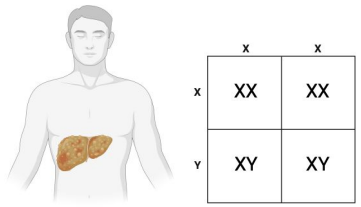
[1]Smedley et al. 2013 *Database*

IMPC
International Mouse Phenotyping Consortium

Queen Mary University of London

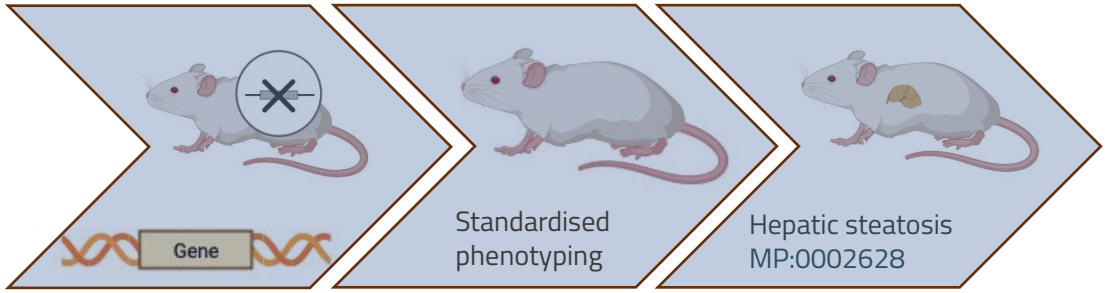
Phenodigm

Phenotyping pipeline



	x	x
x	XX	XX
Y	XY	XY

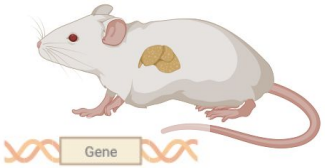
Mendelian diseases



PhenoDigm

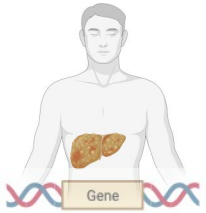
- Novel disease models
- Novel gene discovery

Phenotype matching using orthologous human disease genes



Hepatic steatosis
MP:0002628

MGI



Hepatic steatosis
HP:0001397

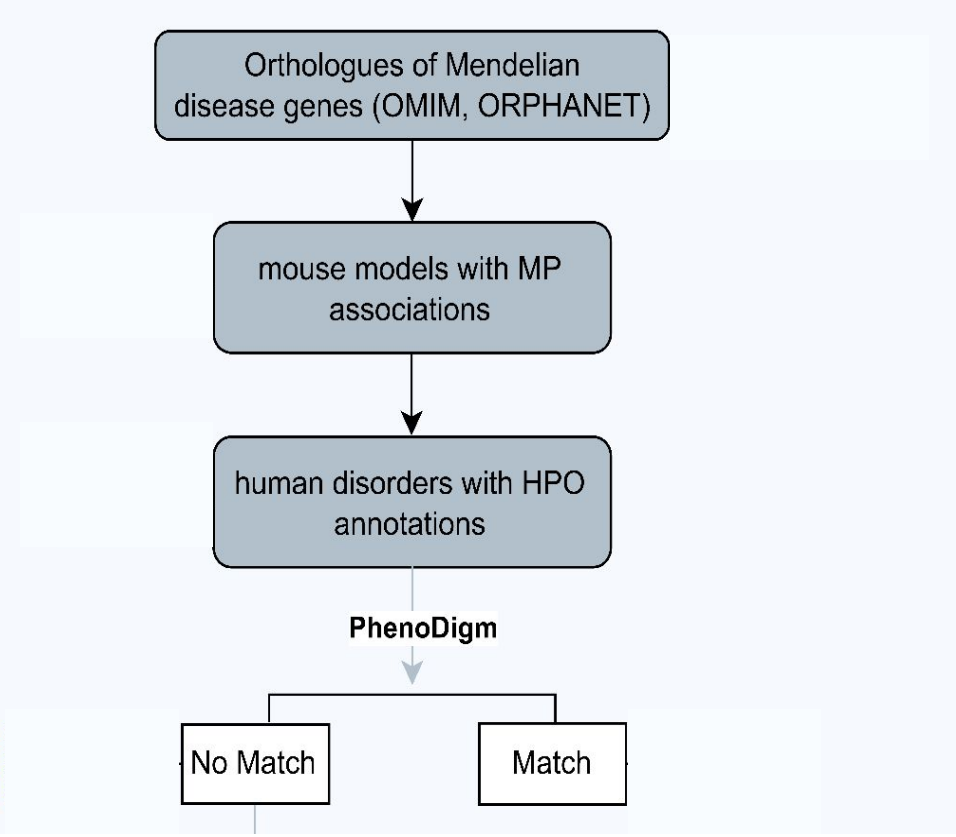
human phenotype ontology

→ **PhenoDigm score** →

Fatty liver disease model

MPI2 Creating a comprehensive catalogue of mammalian gene function.

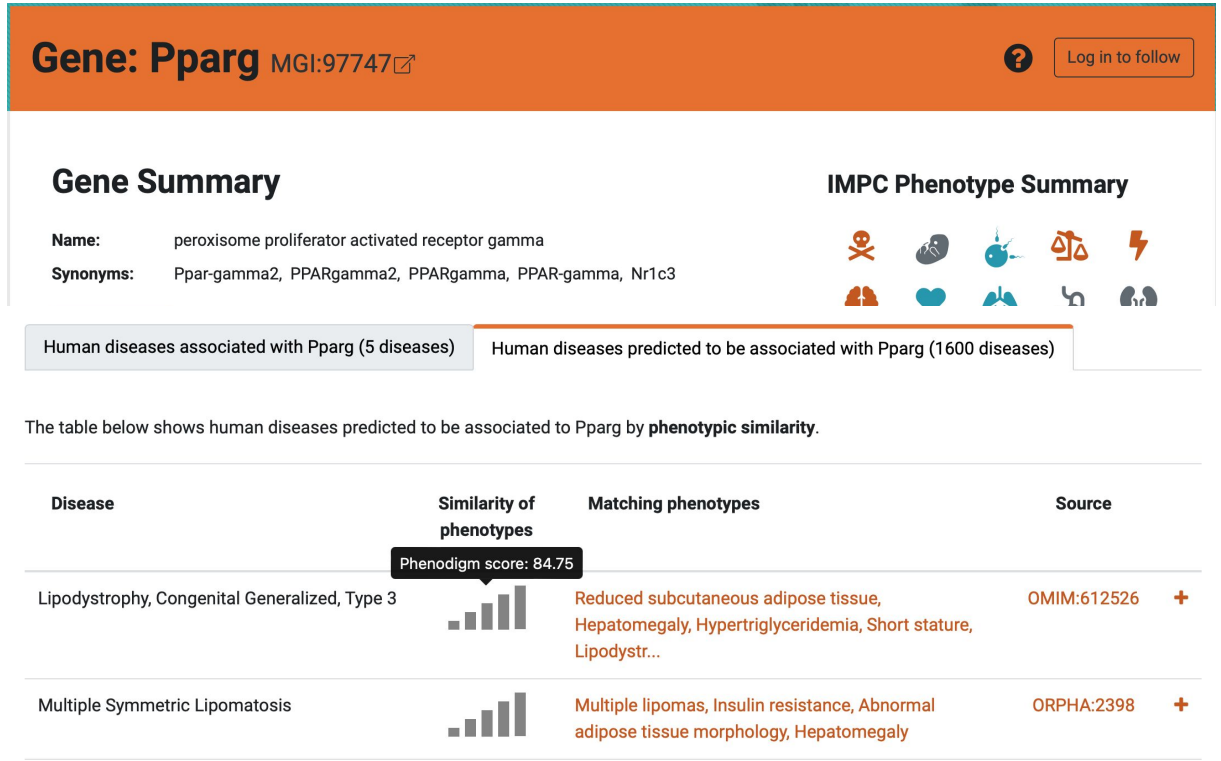
Where does the data come from?



PhenoDigm implementation within IMPC

Website

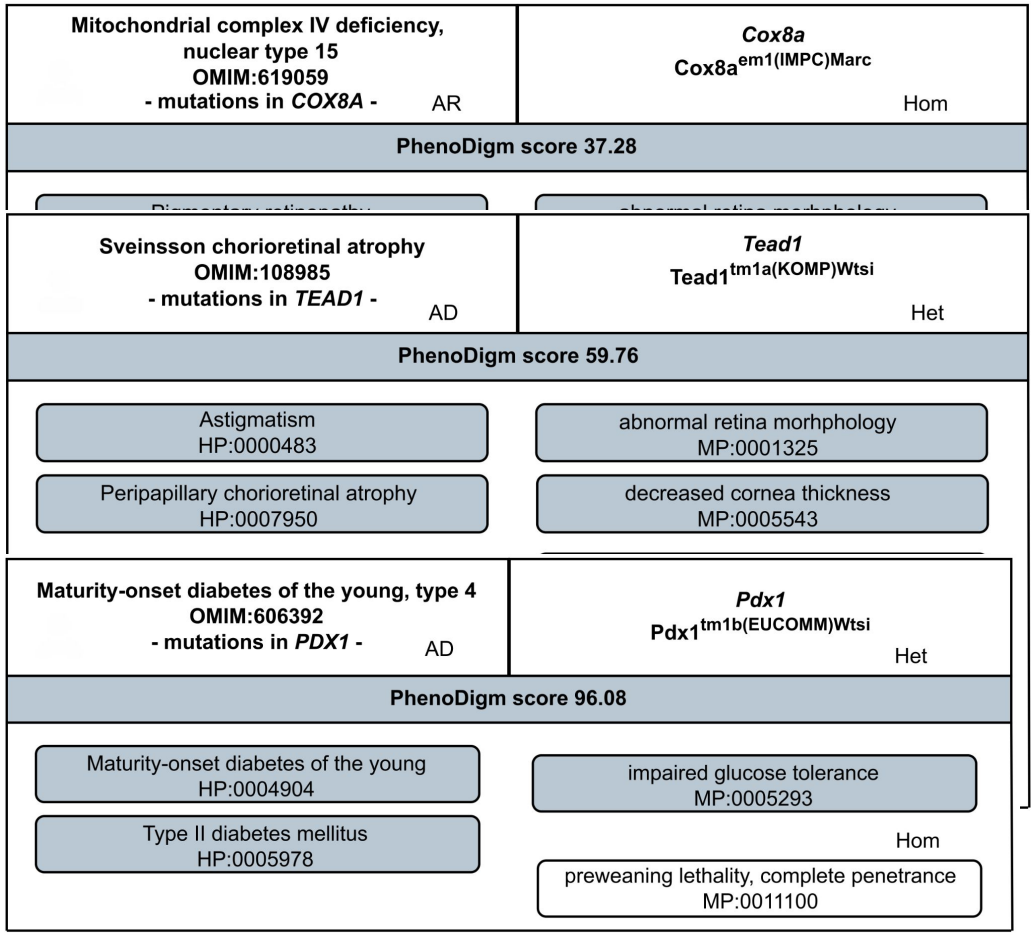
- **Left**: OMIM or Orphanet annotated gene-disease associations
- **Right**: No gene-disease association in OMIM or Orphanet



- Additional resource - Disease Models Portal: https://diseasemodels.research.its.qmul.ac.uk

https://www.omim.org https://www.orpha.net

PhenoDigm scoring within IMPC and examples



Phenodigm core

Information available in the core: *type*

- **disease_model_summary**
- gene
- ontology_ontology
- mouse_model
- ontology
- gene_gene
- disease
- disease_search
- disease_gene_summary

- **Types** are categories/subsets within the core

1. Core overview

- Querying the **phenodigm core**
- Tips:
  - Use keyword **type**
  - Single quotation marks


```
{
  "type":"disease_model_summary",
  "disease_id":"OMIM:227645",
  "disease_term":"Fanconi Anemia, Complementation Group C",
  "model_id":"MGI:3696114",
  "model_source":"MGI",
  "model_description":"C1galt1<ptl1>/C1galt1<ptl1>",
  "model_genetic_background":"C57BL/6-C1galt1<ptl1>",
  "marker_id":"MGI:2151071",
  "marker_symbol":"C1galt1",
  "marker_locus":"6:7845224-7872042",
  "marker_num_models":6,
  "disease_model_avg_raw":0.64,
  "disease_model_avg_norm":28.23,
  "disease_model_max_raw":2.32,
  "disease_model_max_norm":73.31,
  "association_curated":false,
  "disease_matched_phenotypes":["HP:0001903 Anemia",
    "HP:0000086 Ectopic kidney",
    "HP:0000081 Duplicated collecting system",
    "HP:0001876 Pancytopenia",
    "HP:0001896 Reticulocytopenia",
    "HP:0000104 Renal agenesis",
    "HP:0005528 Bone marrow hypocellularity",
    "HP:0001873 Thrombocytopenia",
    "HP:0001518 Small for gestational age",
    "HP:0000085 Horseshoe kidney",
    "HP:0001909 Leukemia",
    "HP:0001875 Neutropenia"],
  "model_matched_phenotypes":["MP:0003179 thrombocytopenia",
    "MP:0000228 abnormal thrombopoiesis",
    "MP:0002828 abnormal renal glomerular capsule morphology",
    "MP:0001859 kidney inflammation",
    "MP:0004756 abnormal proximal convoluted tubule morphology",
    "MP:0002599 increased mean platelet volume",
    "MP:0002962 increased urine protein level",
    "MP:0011402 renal cast",
    "MP:0004505 decreased renal glomerulus number",
    "MP:0002871 albuminuria",
    "MP:0011423 kidney cortex atrophy",
    "MP:0005325 abnormal renal glomerulus morphology",
    "MP:0001262 decreased body weight",
    "MP:0002743 glomerulonephritis"]},
```


```
num_found, df = solr_request(
    core='phenodigm',
    params={
        'q': 'type:disease_model_summary',
    }
)
```

Your request:
https://www.ebi.ac.uk/mi/impc/solr/phenodigm/select?q=type%3Adisease_model_summary

Number of found documents: 8444376

	type	disease_id	disease_term	model_id	model_source	model_description	model_genetic_background
0	disease_model_summary	OMIM:227645	Fanconi Anemia, Complementation Group C	MGI:3696039	MGI	Fign<fi>/Fign<fi>	mixed
1	disease_model_summary	OMIM:227645	Fanconi Anemia, Complementation Group C	MGI:3696114	MGI	C1galt1<ptl1>/C1galt1<ptl1>	C57BL/6-C1galt1<ptl1>
2	disease_model_summary	OMIM:227645	Fanconi Anemia, Complementation Group C	MGI:3696354	MGI	Ercc2<tm3Jhjh>/Ercc2<tm3Jhjh>	involves: 129P2/OlaHsd * C57BL/6 * FVB
3	disease_model_summary	OMIM:227645	Fanconi Anemia, Complementation Group C	MGI:3696370	MGI	Khdrbs1<tm1Rchd>/Khdrbs1<tm1Rchd>	involves: 129S1/Sv * 129X1/SvJ * C57BL/6
4	disease_model_summary	OMIM:227645	Fanconi Anemia, Complementation Group C	MGI:3696775	MGI	Kat6a<tm1Iki>/Kat6a<tm1Iki>	Not Specified

Disease_model_summary – *fields* | *filters*

- **disease_id**: "OMIM:218330"
- **disease_term**: "Cranioectodermal Dysplasia 1"
- **model_id**: "MGI:3587410"
- **model_source**: "MGI"
- **model_description**: "Idua<tm1Clk>/Idua<tm1Clk>"
- **model_genetic_background**: "involves: 129S1/Sv * 129X1/SvJ"
- **marker_id**: "MGI:96418" ← Mouse gene id
- **marker_symbol**: "Idua"
- **marker_locus**: "5:108808197-108832423"
- **marker_num_models**: 7
- **disease_model_avg_raw**: 0.93
- **disease_model_avg_norm**: 42.36
- **disease_model_max_raw**: 2.32
- **disease_model_max_norm**: 72.57
- **association_curated**: false
- **disease_matched_phenotypes**: [HP terms of disease]
- **model_matched_phenotypes**: [MP terms of model]

Documentation at: https://www.ebi.ac.uk/mi/impc/solrdoc/phenodigm.html

2. Query by fields | filters

- The fields available within the core can be used to filter the queries
- Use the boolean operators (AND/OR)
- In this example we filter for models from IMPC only and gene mouse ID MGI:1929872
- We use identifiers from MGI because they are stable (compared to gene symbols)

```
num_found, df = solr_request(
    core='phenodigm',
    params={
        'q': 'type:disease_model_summary AND model_source:"IMPC" AND marker_id:"MGI:1929872"'
    }
)
```

Disease_model_summary - *fields*

- **disease_id**: "OMIM:218330"
- **disease_term**: "Cranioectodermal Dysplasia 1"
- **model_id**: "MGI:3587410"
- **model_source**: "MGI"
- **model_description**: "Idua<tm1Clk>/Idua<tm1Clk>"
- **model_genetic_background**: "involves: 129S1/Sv * 129X1/SvJ"
- **marker_id**: "MGI:96418"
- **marker_symbol**: "Idua"
- **marker_locus**: "5:108808197-108832423"
- **marker_num_models**: 7
- **disease_model_avg_raw**: 0.93
- **disease_model_avg_norm**: 42.36
- **disease_model_max_raw**: 2.32
- **disease_model_max_norm**: 72.57
- **association_curated**: false
- **disease_matched_phenotypes**: [HP terms of disease]
- **model_matched_phenotypes**: [MP terms of model]

Documentation at: https://www.ebi.ac.uk/mi/impc/solrdoc/phenodigm.html

IMPC

International Mouse Phenotyping Consortium

@impc

Exercises: 1-3

mousephenotype.org

Additional skills with Solr

Disease_model_summary – *fields of interest*

- **disease_id**: "OMIM:218330"
- **disease_term**: "Cranioectodermal Dysplasia 1"
- **model_id**: "MGI:3587410"
- **model_source**: "MGI"
- **model_description**: "Idua<tm1Clk>/Idua<tm1Clk>"
- **model_genetic_background**: "involves: 129S1/Sv * 129X1/SvJ"
- **marker_id**: "MGI:96418"
- **marker_symbol**: "Idua"
- **marker_locus**: "5:108808197-108832423"
- **marker_num_models**: 7
- **disease_model_avg_raw**: 0.93
- **disease_model_avg_norm**: 42.36
- **disease_model_max_raw**: 2.32
- **disease_model_max_norm**: 72.57
- **association_curated**: false
- **disease_matched_phenotypes**: [HP terms of disease]
- **model_matched_phenotypes**: [MP terms of model]

Documentation at: https://www.ebi.ac.uk/mi/impc/solrdoc/phenodigm.html

3. Calculations on the fly

- *Solr* allows us to perform **calculations**
- We create a custom field: 'max_avg_diff'
- Use the *sub()* function to subtract two fields
- Tip: you can use '\' to make your queries more readable

```
num_found, df = solr_request(
    core='phenodigm',
    params={
        'q': 'type:disease_model_summary AND model_source:"IMPC" AND marker_id:"MGI:1914217"',
        'fl': 'disease_id,model_id,\
               max_avg_diff:sub(disease_model_max_norm,disease_model_avg_norm)',
        'rows':5
    }
)
```

```
Your request:
https://www.ebi.ac.uk/mi/impc/solr/phenodigm/select?q=type%3Adisease_model_summary+AND+model_source%3A%22IMPC%22+AND+marker_id%3A%22MGI%3A1914217%22&fl=disease_id%2Cmodel_id%2C+++++++++++++++++max_avg_diff%3Asub%28disease_model_max_norm%2Cdisease_model_avg_norm%29&rows=5

Number of found documents: 109
```

	disease_id	model_id	max_avg_diff
0	OMIM:229200	MGI:6162234#het#early	64.270004
1	OMIM:250500	MGI:6162234#het#early	27.779995
2	OMIM:241520	MGI:6162234#het#early	25.109997
3	OMIM:242840	MGI:6162234#het#early	67.130005
4	OMIM:155950	MGI:6162234#het#early	15.110001

4. Sorting

- Sort existing fields using the *sort* parameter:
  - 'sort': field *asc* or *desc*
  - 'sort': 'disease_term desc'
- Sort on values calculated on the fly

```
num_found, df = solr_request(
    core='phenodigm',
    params={
        'q': 'type:disease_model_summary AND model_source:"IMPC" AND marker_id:"MGI:1914217"\


        'fl': 'disease_id,model_id\
        max_avg_diff:sub(disease_model_max_norm,disease_model_avg_norm)',
        'sort':'sub(disease_model_max_norm,disease_model_avg_norm) asc'
    }
)
```

Your request:
https://www.ebi.ac.uk/mi/impc/solr/phenodigm/select?q=type%3Adisease_model_summary+AND+model_source%3A%22IMPC%22+AND+marker_id%3A%22MGI%3A1914217%22++++++++++++++++AND+disease_model_max_norm%3A%5B%2A+TO+%2A%5D++++++++++++++++AND+model_matched_phenotypes%3A+%5B%2A+TO+%2A%5D&fl=disease_id%2Cmodel_id%2C++++++++++++++++max_avg_diff%3Asub%28disease_model_max_norm%2Cdisease_model_avg_norm%29&sort=sub%28disease_model_max_norm%2Cdisease_model_avg_norm%29+asc&rows=5

	disease_id	model_id	max_avg_diff
0	ORPHA:1410	MGI:6162234#het#early	12.299999
1	OMIM:259690	MGI:6162234#het#early	14.300003
2	OMIM:155950	MGI:6162234#het#early	15.110001
3	OMIM:166260	MGI:6162234#het#early	20.909996
4	OMIM:278150	MGI:6162234#het#early	21.299995

Disease_model_summary – *fields of interest*

- **disease_id**: "OMIM:218330"
- **disease_term**: "Cranioectodermal Dysplasia 1"
- **model_id**: "MGI:3587410"
- **model_source**: "MGI"
- **model_description**: "Idua<tm1Clk>/Idua<tm1Clk>"
- **model_genetic_background**: "involves: 129S1/Sv * 129X1/SvJ"
- **marker_id**: "MGI:96418"
- **marker_symbol**: "Idua"
- **marker_locus**: "5:108808197-108832423"
- **marker_num_models**: 7
- **disease_model_avg_raw**: 0.93
- **disease_model_avg_norm**: 42.36
- **disease_model_max_raw**: 2.32
- **disease_model_max_norm**: 72.57
- **association_curated**: false
- **disease_matched_phenotypes**: [HP terms of disease]
- **model_matched_phenotypes**: [MP terms of model]

Documentation at: https://www.ebi.ac.uk/mi/impc/solrdoc/phenodigm.html

Calculating Phenodigm score

- $$Phenodigm\ score\ =\ AVG(\boldsymbol{disease\_model}\_\max\boldsymbol{\_norm}, \boldsymbol{disease\_model\_avg\_norm})$$

IMPC

International Mouse Phenotyping Consortium

@impc

Advanced Exercises: 4-6

mousephenotype.org

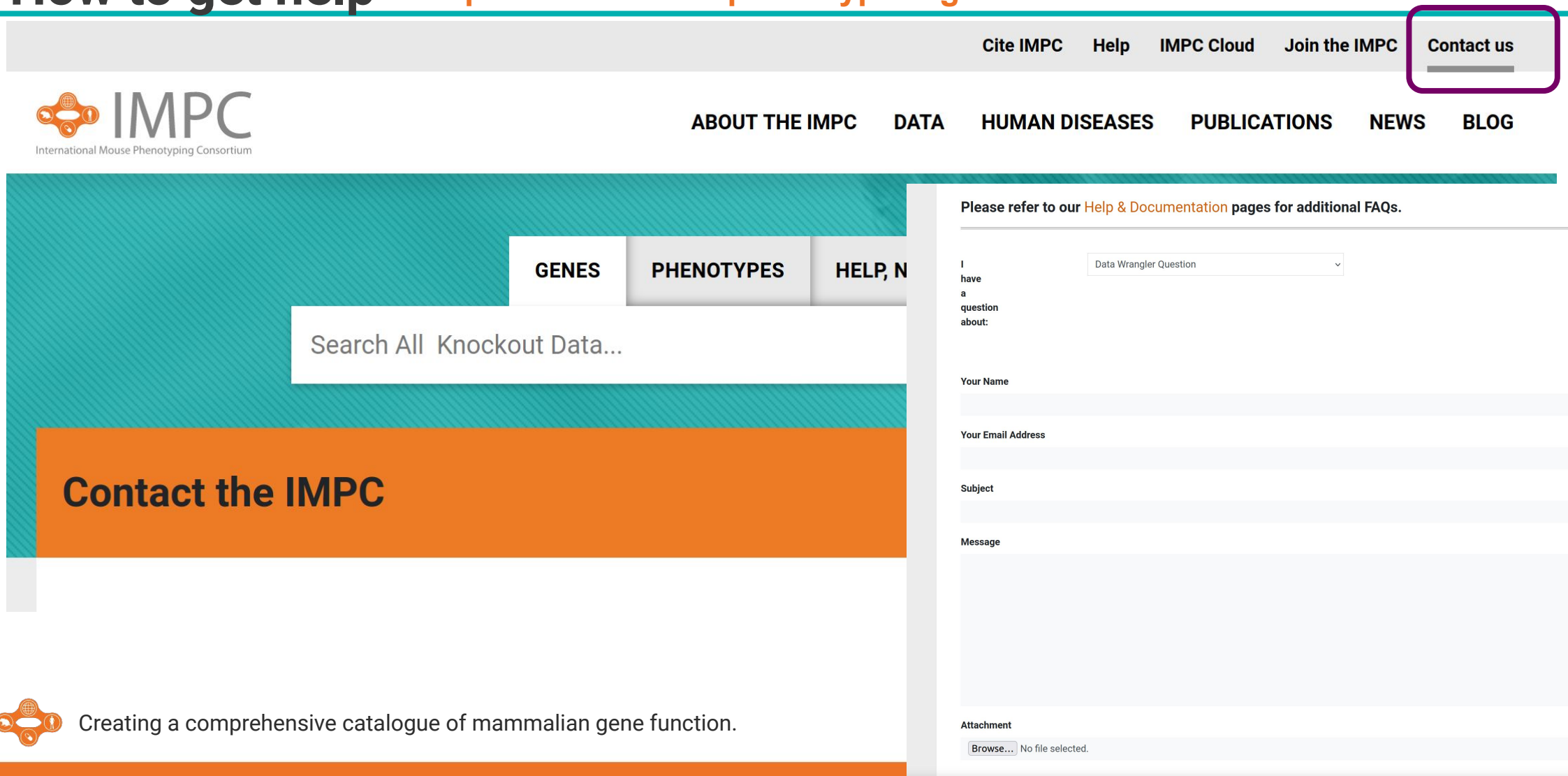
Important information

1. The JupyterHub service will be available for a week after the workshop to allow you to review the exercises after the workshop

2. To access the presentations and the Jupyter notebook remember to bookmark: https://github.com/mpi2/impc-data-api-workshop

How to get help

https://www.mousephenotype.org/contact-us/



Creating a comprehensive catalogue of mammalian gene function.