PhenoMeNal: accessing metabolomics workflows in Galaxy

PhenoMeNal (Phenome and Metabolome Analysis) is a standardised e-infrastructure that supports the data processing and analysis pipelines for molecular phenotype data generated by metabolomics applications. This course will give you an overview of PhenoMeNal, how to create your cloud research environment, and how to access Galaxy workflows for metabolomics data.

Learning objectives:

- Describe what PhenoMeNal is and when to use it
- Create your cloud research environment
- Access Galaxy workflows via PhenoMeNal

What is PhenoMeNal

A phenotype [5] is a set of observable characteristics of an organism and is the combined result of the organism's genetic makeup and external environmental factors. Metabolic phenotyping, or measuring and modelling all metabolites of an individual, provides insights into disease risk factors and aetiology. This information can be used for personalised medicine (1 [6]). PhenoMeNal is a comprehensive and standardised e-infrastructure that supports data processing and analysis workflows for molecular phenotype data generated by metabolomics studies. This will allow us to improve our ability to monitor health and to detect, treat and manage diseases. The infrastructure is generic and can also be used for other omics studies and non-clinical applications if needed but may require integration of tools from those domains.

PhenoMeNal Tools and Workflows

PhenoMeNal provides access to popular metabolomics analysis tools such as XCMS [7], rNMR [8], IPO [9], Workflow4Metabolomics [10] and MetFrag [11] among many others, and combines them with the computing power of the cloud environment [12]. All of these are available through user friendly workflow environment tools. You can create workflows and browse metabolomics tools from the PhenoMeNal Gateway [4] homepage (Figure 1).
Learn more about cloud environments on Wikipedia [12].

Figure 1 The PhenoMeNal Gateway homepage is easily accessible by clicking on the 'Portal' tab in the menu on the PhenoMeNal project website [13]. You can create workflows and browse metabolomics tools from this page.

Why do we need PhenoMeNal?

When doing computational data analysis, researchers will often be confronted with the difficult reality of having to install complex scientific software and tools. In addition, you must set up intricate methods of connecting these tools, all while dealing with data standard requirements and data transformations. This typically requires a fairly high level of expertise in bioinformatics, which is typically not accessible to most researchers. PhenoMeNal aims to bridge this technical gap by providing an automatic way of using and connecting a plethora of metabolomics software and tools.

Real challenges and PhenoMeNal solutions
Caroline is a clinician-scientist who is profiling the metabolic phenotype of people with bladder cancer. She needs a way to analyse her data so that she can identify biomarkers that can be used to detect bladder cancer. She has little programming and computational expertise.

**Challenges**

**No programming experience**

*PhenoMeNal solution*

PhenoMeNal allows someone with little programming experience to put metabolomics tools together in well-defined pipelines, by using a graphical workflow environment called Galaxy. Galaxy has been successfully used in other fields of life sciences for more than a decade.

**Complicated computer requirements**

*PhenoMeNal solution*

PhenoMeNal allows you to access the tools in an online environment where the complexity of installing tools and setting up a computer cluster, networks and security considerations have already been taken care of. In essence, you get your own cluster with the software you need preinstalled to do metabolomics data analysis.

**Extreme data volumes and data transfers**

*PhenoMeNal solution*

People often take their data to a collaborator’s computer cluster for data analysis, where all tools have been previously installed. In a world where datasets are becoming larger and larger, moving data around becomes more and more complex. PhenoMeNal reverses this paradigm, by allowing users to access the computer cluster from where the data is.

**Scalability**

*PhenoMeNal solution*

Traditionally, the speed of data analysis is limited by the availability, number and power of your local computing machinery. By using public cloud providers, like Amazon AWS or Google GCE, PhenoMeNal provides the ability to run data analysis in as few, or as many machines as necessary to get your analysis done as quickly as you need it.
How PhenoMeNal can help with your research

When to use PhenoMeNal

You can use PhenoMeNal:

- To analyse your metabolomics data using tools from the PhenoMeNal App Library within the Galaxy workflow environment.
- To enable scalable and reproducible research.

The PhenoMeNal App Library [16] is a catalogue of all the tools that you can expect to find in PhenoMeNal. In contrast to some of the app stores that you might be familiar with, you don't need to do anything to use a particular tool in PhenoMeNal. The tools are automatically available.

PhenoMeNal and reproducible research. Each release of PhenoMeNal is associated with defined versions of the underlying tools. This means that your results are comparable as long as you use the same PhenoMeNal release each time. Additionally, the use of Galaxy workflows allows parameters and settings to be easily saved and reproducibly applied.

When not to use PhenoMeNal

PhenoMeNal cannot be used to access cloud services for purposes other than metabolomics data analysis. For this, you need to contact the cloud providers for further information.

Limitations:

- When using the free, PhenoMeNal-provided cloud, you can only upload a few gigabytes of data. You can upload more data by using PhenoMeNal in combination with a commercial cloud provider (e.g., Amazon) for which you will need to buy credit, or with your institution’s OpenStack installation (a private cloud provider).

- If an app is not in the App Library, then you won't be able to use it as part of a Galaxy workflow in PhenoMeNal. With some joint effort, we can make new tools available and we provide extensive documentation for external developers [17] to migrate new tools to the PhenoMeNal architecture.

- Non-standard data formats might require some additional data transformation [18] steps before they can be uploaded in PhenoMeNal.

Creating a Cloud Research Environment in PhenoMeNal

You can create a cloud research environment (CRE) in PhenoMeNal from the PhenoMeNal Gateway homepage [4] by clicking 'Create Cloud Research Environment' button or the 'CRE' tab (Figure 2A).
From this page you can also:

- Create a Galaxy workflow using the 'Galaxy Workflow' button (Figure 2B).
- Browse the App library [16] for free metabolomics data analysis tools (Figure 2C).
- Access tutorials and documentation [19] (Figure 2D).
- Access the PhenoMeNal CRE using Jupyter Ipython (not discussed in this tutorial) (Figure 2E).

**Figure 2** The PhenoMeNal Gateway homepage.

Creating a PhenoMeNal CRE takes three steps(Figure 3):

1. Sign into the PhenoMeNal Gateway by using the Elixir Single Sign-On (SSO).
2. Choose the cloud provider you want to use and sign in. You can either use the freely provided PhenoMeNal cloud or a commercial cloud provider such as Amazon or Google.
3. Sign into Galaxy within the PhenoMeNal CRE.

Make sure you keep track of your account details as you will need them each time you access the PhenoMeNal CRE.

You can find step-by-step instructions on creating a CRE in PhenoMeNal in these PDF guides:

- Creating a PhenoMeNal CRE with the free PhenoMeNal cloud [20]
- Creating a PhenoMeNal CRE with Amazon web services cloud [21]
Finding your way around Galaxy

Once you have created your cloud research environment in PhenoMeNal, you will be able to access the PhenoMeNal Galaxy instance which contains a set of tools for metabolomics analysis.

You will first need to log into the PhenoMeNal Galaxy instance, with the username and password you created in the previous steps (see the detailed instructions in the PDFs on the previous page).

Go to the ‘User’ drop-down menu (Figure 4) and select the ‘Login’ option.

Note that the ‘Workflow’ and ‘Visualization’ drop-down menus are disabled before login.

The way that you use and interact with Galaxy through PhenoMeNal is the same as if you were to
use it via the Galaxy website. The advantage being that coupling Galaxy with the cloud gives you increased computing power to perform your analyses.

To familiarise yourself with the Galaxy interface we recommend these guided tours [22] from the Galaxy team. After you go through these tours, you should feel comfortable starting to use our tools in the PhenoMeNal Galaxy instance.

Some of the basic features are highlighted in the figure below (Figure 5).

Note that the available features vary depending on whether you are using the free PhenoMeNal cloud or a commercial cloud provider.

Figure 5 Key menus and features of the PhenoMeNal Galaxy instance.

How to use Galaxy: guided example

This video gives you a taster of some of the things that can be done in Galaxy. It shows you how to explore, load, import, view and visualise an existing dataset.

Getting help and support on PhenoMeNal

We always welcome your feedback and questions regarding PhenoMeNal or this tutorial.
Simply email us or click the feedback button on the PhenoMeNal help pages [19].

On the help pages you will find links to tutorials and documentation as well as a glossary [23] of technical terms used in the PhenoMeNal project. There are also more videos and screencasts on our YouTube channel [24].

Figure 6 PhenoMeNal help page.

Summary

What is PhenoMeNal?

PhenoMeNal is a comprehensive and standardised e-infrastructure that supports the data processing and analysis pipelines for molecular phenotype data generated by metabolomics applications.

Why use PhenoMeNal?

The accumulation of unprecedented numbers of high dimensional and increasingly complex datasets covering metabolic phenotyping, personal genomes and other data, offers great promise for capturing subtle and complex biologically imprinted effect of accumulated exposure (exposome) on health. Processing and analysis of such data, however, requires advanced computation and data management. PhenoMeNal combines the power of cloud computing with easy to use data analysis workflows to overcome these challenges.

What tools are provided?
PhenoMeNal provides a wide range of metabolomics tools (29 tools and counting). To browse the available tools visit the App Library [16].

**How to access the PhenoMeNal cloud research environment**

PhenoMeNal cloud research environment can be accessed via the PhenoMeNal Gateway [25]. All you need is an internet connection.

**How to use Galaxy**

The way that you use and interact with Galaxy through PhenoMeNal is the same as if you were to use it via the Galaxy website. To familiarise yourself with the Galaxy interface we recommend these guided tours [22] from the Galaxy team.

**Your feedback**

Please tell us what you thought about this course. Your feedback is invaluable and helps us improve our courses and enhance your learning experience.

**Learn more**

**Recommended courses**

To learn more about metabolomics in general we recommend our course Metabolomics: An introduction [26].

You might also be interested in the MetaboLights: Quick tour, [27]

PhenoMeNal uses Galaxy to create metabolomics workflows. To familiarise yourself with Galaxy we recommend these guided tours [22] from the Galaxy team.

**References**

**References**


Contributors

Namrata Kale [34]

EMBL-EBI
Scientific Database Curator - Steinbeck team: Cheminformatics and metabolism

Namrata S. Kale obtained her PhD in natural product chemistry from Indian Institute of Technology, New Delhi, India. She is currently a member of the Cheminformatics and Metabolism group at EMBL-EBI, where she works as a Project Manager for PhenoMeNal.

Pablo Moreno [3]

EMBL-EBI
Senior Bioinformatician - Steinbeck team: Metabolomics and Molecular Informatics

Pablo is the Senior Bioinformatician for the PhenoMeNal project, within the Cheminformatics & Metabolism group at EMBL-EBI. After his PhD and Post-doc in Bioinformatics with Christoph Steinbeck at EMBL-EBI, which dealt with the inference of complete Metabolomes based on multiple sources of data and various aspects related to metabolism representation, Pablo was appointed to lead the
Bioinformatics Core facility at the Cambridge Institute for Medical Research (CIMR), dependent of the University of Cambridge. Two years later, Pablo came back to EMBL-EBI to work on PhenoMeNal.

Sijin He [2]

EMBL-EBI
Staff Member - Steinbeck team: Metabolomics and Molecular Informatics

Dr. Sijin He received a BSc in Mathematics (1st Class Hon) from University College London in 2007 and a MSc in Computing Science from Imperial College London in 2008. In August 2013, he received PhD in Computer Science at Imperial College London on cloud computing. He then worked as post-doctoral research associate at Data Science Institute at Imperial College London from 2013 to 2016. He is now working as a full stack software developer at EMBL-EBI.

Source URL: http://www.ebi.ac.uk/training/online/course/phenomenal-accessing-metabolomics-workflows-galaxy

Links
1. http://www.ebi.ac.uk/training/online/trainers/Namrata%20Kale
2. http://www.ebi.ac.uk/training/online/trainers/sijin_11005
3. http://www.ebi.ac.uk/training/online/trainers/pmoreno_11007
5. https://en.wikipedia.org/wiki/Phenotype
8. http://rnmr.nmrfam.wisc.edu/
11. http://msbi.ipb-halle.de/MetFrag/
14. http://www.ebi.ac.uk/training/online/glossary/metabolomics
15. https://galaxyproject.org/
18. http://www.ebi.ac.uk/training/online/glossary/transformation
20. http://www.ebi.ac.uk/training/online/sites/ebi.ac.uk.training.online/files/CreatingaPhenoMeNalCloudResearchEnvironmentwiththeFreePhenoMeNalCloud_0.pdf
22. https://usegalaxy.org/tours

Page 11 of 12
[24] https://www.youtube.com/channel/UCXGAvsVNOQk-aUpckjRC8Ang
[26] http://www.ebi.ac.uk/training/online/course/introduction-metabolomics
[27] http://www.ebi.ac.uk/training/online/course/metabolights-quick-tour-0
[29] http://europepmc.org/articles/PMC4648992
[33] http://pubs.acs.org/doi/abs/10.1021/acs.jproteome.6b00125
[34] http://www.ebi.ac.uk/training/online/trainers/Namrata Kale