PhenoMeNal Gateway: Metabolomics data analysis in the cloud

Namrata Kale [1], Sijin He [2], Pablo Moreno [3]

- Chemical biology
- Cross-domain
- Beginner
- 1 hour

The PhenoMeNal Gateway is the portal to your own PhenoMeNal Cloud Research Environment (CRE) containing an array of essential metabolomics applications, all available through a user-friendly Galaxy workflow environment. This course will give you an overview of PhenoMeNal Gateway [4], how to create your cloud research environment, and how to use the PhenoMeNal Galaxy instance for workflows for metabolomics data.

Learning objectives:

- Describe what PhenoMeNal is and when to use it
- Create your cloud research environment
- Using the PhenoMeNal Galaxy instance

What is PhenoMeNal?

PhenoMeNal infrastructure

PhenoMeNal (Pheno and Metabolome aNalysis) is a comprehensive, scalable, standardised, secure, and sustainable e-infrastructure that enables metabolomics [5] researchers to process and analyse both small and large-scale metabolomics datasets. The e-infrastructure, funded by the European Commission's H2020 programme (grant agreement number: 654241), offers a solution to the challenges being faced by those handling, processing and analysing complex and large-scale metabolomics datasets. The computational services provided within PhenoMeNal enables you to further improve the understanding of the causes and mechanisms underlying health, healthy ageing and disease. 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.

A phenotype [6] 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 [7]).
PhenoMeNal Gateway: portal to metabolomics tools and workflows

The PhenoMeNal infrastructure provides access to popular metabolomics analysis tools such as XCMS [8], rNMR [9], IPO [10], Workflow4Metabolomics [11] and MetFrag [12] among many others (see the App library for a full list), and combines them with the computing power of the cloud environment [13]. All of these are available through user friendly workflow environment tools. You can create your cloud research environment (CRE) and browse metabolomics tools from the PhenoMeNal Gateway [14] homepage (Figure 1).

Learn more about cloud environments on Wikipedia [13].

Figure 1 The PhenoMeNal Gateway homepage is easily accessible by clicking on the 'Portal' tab in the menu on the PhenoMeNal project website [15]. You can create your cloud research environment (CRE) 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 [16] 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

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 [21] 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 [22] to migrate new tools to the PhenoMeNal architecture.

- Non-standard data formats might require some additional data transformation [23] steps before they can be uploaded in PhenoMeNal.

How to use PhenoMeNal Gateway?

The PhenoMeNal Gateway homepage provides you access to resources to create your own cloud research environment (CRE) using cloud provider of your choice (e.g. AWS, GCP), test drive our free public instance, browsing the service catalogue of metabolomics [5] tools and online training.

Our focus on clinical metabolomics ensures that we take the utmost care in securing your environment and data. The Cloud Research Environment adheres to the strictest security measures. At no point do we store any information related to your own cloud provider.

Navigating the PhenoMeNal Gateway homepage

How to create your own CRE?

You can create a cloud research environment (CRE) in PhenoMeNal from the PhenoMeNal Gateway homepage [14] by clicking ‘Create Cloud Research Environment’ button (A) or the ‘CRE’ tab (B) or the ‘Sign in’ tab (C) on the top menu (Figure 6).
Creating a PhenoMeNal CRE takes four steps (Figure 7)

1. Sign into the PhenoMeNal Gateway by using the Elixir Single AAI (A).
2. Choose your preferred login account: Google, Institutional account, Orcid or Linkedin (B).
3. 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 or Open Stack (C).
4. Register you account (D).

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 on our help [28] pages:

Starting a PhenoMeNal CRE via PhenoMeNal Portal [29]

![Figure 7 Registering with PhenoMeNal](image)

**Finding your way around Galaxy**

Once you have created your cloud research environment, you will be able to access the PhenoMeNal Galaxy instance which contains a set of tools for metabolomics analysis. Use your username and password created in the previous steps to log into the PhenoMeNal Galaxy instance (Figure 8).
Navigating the PhenoMeNal Galaxy instance homepage

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 [30]. After you go through these tours, you should feel comfortable starting to use our tools in the PhenoMeNal Galaxy instance.

The PhenoMeNal Galaxy instance home page consists of a top menu (A), ‘Tools’ panel (B) and ‘History’ panel (C). The homepage also provides links to workflow tutorials (D) and user documentations under ‘Support’ (E) (Figure 9).

Figure 9 Top menu and features of the PhenoMeNal Galaxy instance.

PhenoMeNal Galaxy homepage top menu allow you to:

- ‘Analyse data’: Set tools parameters
- ‘Workflow’: access stored, shared workflows
- ‘Shared Data’: access data libraries, histories, workflows
- ‘Visualization’: access new track browser, saved visualisation, interactive environments
'Help': for user support, mailing lists, videos, wiki, citation, interactive tours
'User': user management
'Scratch book': change the way data is shown

How to use PhenoMeNal Galaxy Instance?

The PhenoMeNal Galaxy instance is free for all users and does not require pre-existing cloud provider credentials (e.g. AWS or GCP credentials)

Importing and visualising existing datasets in PhenoMeNal Galaxy Instance (guided example).

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 (Figure 13).

On the help pages you will find links to tutorials and documentation as well as a glossary of technical terms used in the PhenoMeNal project. There are also more videos and screencasts on our YouTube channel.
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 wide range of metabolomics tools (29 tools and counting). To browse the available tools visit the App Library [21].

How to access the PhenoMeNal cloud research environment

PhenoMeNal cloud research environment can be accessed via the PhenoMeNal Gateway [37]. 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 [38] 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

For more detailed information on PhenoMeNal infrastructure, tools and workflows, we recommend watching our individual webinars:

- PhenoMeNal: Metabolomics data analysis in the cloud [39]
- Challenges and Opportunities with Virtual Research Environments [40]
- LC/MS data analysis with XCMS and MetFrag on PhenoMeNal [41]
- Metabolomics in the cloud: scaling computational tools to big data [42]
- Statistical Workflows in PhenoMeNal [43]

To learn more about metabolomics in general we recommend our course Metabolomics: An introduction [44].

You might also be interested in the MetaboLights: Quick tour [45]

PhenoMeNal uses Galaxy to create metabolomics workflows. To familiarise yourself with Galaxy we recommend these guided tours [38] from the Galaxy team.

References

References

Other selected reading


Contributors

Namrata Kale [52]
EMBL-EBI
Project Manager - O'Donovan team: Metabolomics

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 Metabolomics 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: https://www.ebi.ac.uk/training/online/course/phenomenal-gateway-metabolomics-data-analysis-cloud

Links
[1] https://www.ebi.ac.uk/training/online/trainers/Namrata%20Kale
[2] https://www.ebi.ac.uk/training/online/trainers/sijin_11005
[3] https://www.ebi.ac.uk/training/online/trainers/pmoreno_11007
[5] https://www.ebi.ac.uk/training/online/glossary/metabolomics
[7] https://www.ebi.ac.uk/training/online/course/phenomenal-accessing-metabolomics-workflows-galaxy/references
[16] https://www.ebi.ac.uk/training/online/glossary/phenotype