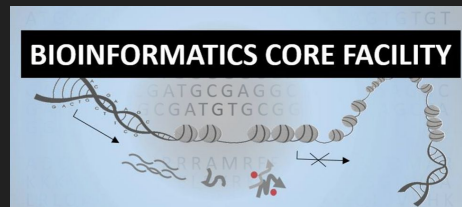


Welcome to ABC.11

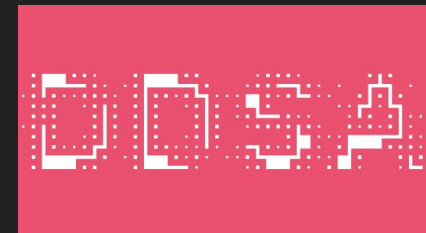
16. January 2025

<https://abc.au.dk>

abcafe@au.dk



Health
Data Science
Sandbox



Danish Data
Science
Academy

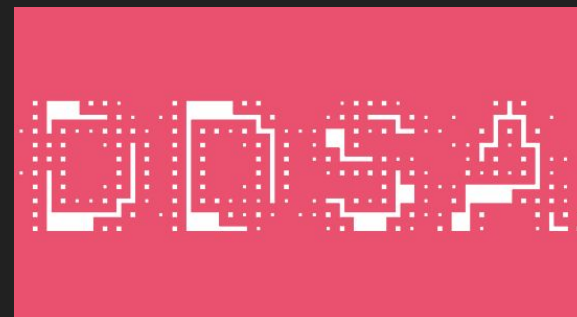
Agenda

- What's new
- Interesting things from you?
- Topic presentation
- Tutorial and/or open coding

What's new

Some updates

- Next cafe dates
 - Jan 30
 - Feb 6
 - Feb 27
 - Mar 13
- You will always get an invitation if you are on the newsletter
- GenomeDK workshop Series
 - Feb 26
 - Apr 10
 - Jun 4
- Funding from DDSA



What's new

Starting a resources page

Aiming to:

databases

code snippets

useful courses

open-source stuff

 ABC - Accessible Bioinformatics Cafe

[Home](#) [Calendar](#) [Documentation](#) [Resources ▾](#)

[Databases](#)

[Bash tips](#)

Archive of OMICS databases with open datasets

[OMICS](#) [DATABASES](#)

A searchable table of periodically updated and daily tested OMICS databases with open access datasets.

AUTHOR

PUBLISHED

Samuele Soraggi, Marie Sønderstrup

December 9, 2024

In this notebook, we provide a searchable table of periodically updated and daily tested OMICS databases with open access datasets. The table is based on multiple online sources and inputs from our colleagues and ABC participants. The links are tested automatically every day at midnight, and an error in opening the related webpage is signaled in the table. When we verify the malfunctioning of the link, we remove the database from the list after a reasonable grace period.

Note

For any contribution or suggestion, please contact us at [abcafe \[at\] au.dk](mailto:abcafe[at]au.dk), or open an issue on the Github repository of this webpage. Any input is considered, especially if you produced your own database and want to share it with the community!!!

[Copy](#) [CSV](#) [Excel](#)

Gene Expression Omnibus (GEO)	Public functional genomics data repository
Sequence Read Archive (SRA)	Raw sequencing data and alignment information from GEO data
Human Brain Transcriptome	Transcriptome data and associated metadata for the developing
Allen brain atlas	A wide set of tools and data focused on the brain
BrainSpan	Foundational resource for studying transcriptional mechanisms
EnrichR	Website/interactive and collaborative gene list enrichment tool
SysNDD	Gene disease relationships in neurodevelopmental disorders (N
Adult Genotype Tissue Expression (GTEx)	Tissue and cell-specific gene expression and regulation across in
Human protein atlas	The open access resource for human proteins
FANTOM	Functional elements in mammalian genomes

Showing 1 to 10 of 11 entries

1

2

Interesting things from you?

Running Language Models Locally



```
curl http://localhost:11434/v1/chat/completions \
-H 'Content-Type: application/json' \
-d '{
  "model": "llama2",
  "messages": [
    {
      "role": "user",
      "content": "Hello!"
    }
  ]
}'
```

LMs, Ollama & Requirements



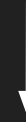
Get up and running with large language models.

- A language model is a type of AI developed to understand, create, and predict human language.
- Ollama is a platform for running language models (LMs) locally—**offline use, ensures data privacy**.
 - Ollama allows you to run a variety of language models (small and large).
- Requirements: a CPU from 2015 onwards & 8 GB RAM will be OK—for the small models.

Setting up Ollama

1. Download & install:

ollama-setup.exe  ollama.dmg 
<https://ollama.com/>



2. Install and run a model:

Open the terminal and type:

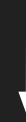
``ollama run llama3.2``

This command will install and execute the model when it's installed

```
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS C:\Users\manpe> ollama run llama3.2
>>> Send a message (/? for help)
```



3. Enter your prompts

'Correct the grammar and punctuation of this paragraph'

Why Small Language Models Make Sense.

- Running a large language model is (very) computationally intensive.
- Small models (3–7B parameters) need only 8GB RAM (vs 32GB+ for larger models)
- They need less space in the hard drive (2–4GB vs 40GB+ for larger models)
- They can run efficiently without dedicated GPUs

llama3.2

Meta's Llama 3.2 goes small with 1B and 3B models.

tools

1b

3b

↓ 6.7M Pulls ⌚ Updated 3 months ago

llama3.2 is an example of a small language model—but there are more.

When they don't make as much sense

More prone to errors (hallucinations)

Limited contextual understanding

Struggles with complex reasoning

May not follow instructions as reliably

Complex tasks:

- Advanced reasoning
- Mathematical problem-solving
- Specialized domain knowledge

Larger models (e.g., 7b, 13b, 175b...) have more parameters and produce higher-quality outputs—but require way more computational resources.

‘Soon’

Setting up ollama large models in Genome DK
(First we need to figure out how to) ͇_(ツ)͇

Tutorial and open coding

At our home <https://abc.au.dk/Documentation> you can find

- Today's tutorial on single cell preprocessing and data structure with Seurat
- Previous tutorials + 1 conference workshop

Or you can code and ask for help or generic coding/bioinf/data science questions



And have cake