

AI TOOLS & BEST PRACTICES FOR RESEARCHERS

Bringing AI into your workspace

Session 1 · VS Code + GitHub Copilot on your local machine

CCR Genetics Branch
June 23, 2026



Motivation: bring AI to your project

For many users, AI is synonymous with web chatbots. These tools are useful for research, but analyzing data requires copying data files and code in and out of the interface.



Lost context

A chatbot cannot view other files, folder layouts, or how your data is shaped.



Data security concerns

Pasting sensitive data into an app can be a governance and security problem.



Lack of reproducibility

Without the underlying code, analyses can't be reproduced.



Instead: bring AI to your workspace

Copilot runs in VS Code on your local file system or Biowulf where your data already lives. It reads your files, writes scripts, and runs them in place. Nothing gets uploaded anywhere.

Agentic AI: “an artificial intelligence system that can accomplish a specific goal with limited supervision”

Agenda

1



Safety & data security

How to keep your files safe and analyses reproducible.

2



Set up VS Code

Launch a VS Code session on your local machine.

3



Connect GitHub Copilot

Link your GitHub account and sign in to Copilot inside VS Code.

4



Install the Copilot Chat extension

Add the extension.

5



Understand Models

Know which models are available and how requests / tokens are billed.

6



Hands-on: build a volcano plot

Prompt Copilot to write and run a plotting script on a DESeq2 table.

Use AI with caution

AI is a fast assistant, not an oracle. It's a “stochastic parrot” trained on GitHub and Stack Overflow — it predicts plausible code, it doesn't understand your biology.

NEVER

- ✗ **Expose PHI or PII to an AI tool***
- ✗ **Blindly accept AI-generated code**
- ✗ **Let AI agents run destructive commands**

ALWAYS

- ✓ **Review commands before running**
- ✓ **Check every script and output**
- ✓ **Watch for silent assumptions**

***NIH enterprise AI tools are approved for PII, but with many caveats**

(<https://nih.sharepoint.com/sites/NIH-ai/SitePages/Responsible-AI.aspx#approved-tools-and-data-types>)

Don't blindly accept AI-generated code



NIH Guidance

<https://nih.sharepoint.com/sites/NIH-ai/SitePages/Responsible-AI.aspx>

Do not rely on the technology to be a software developer by proxy: All well-written code must adhere to security design and ethical principles. All code output needs to be reviewed for completeness, quality, efficiency, and, most of all, security. Leverage manual and automated validation tools and testing technologies to help ensure these factors. If you cannot identify or understand what a piece of AI generated code does, you should not use it.



Treat every output as a draft. AI outputs may be incomplete, inaccurate, biased, or fabricated.

Data protection and reproducibility

Raw data

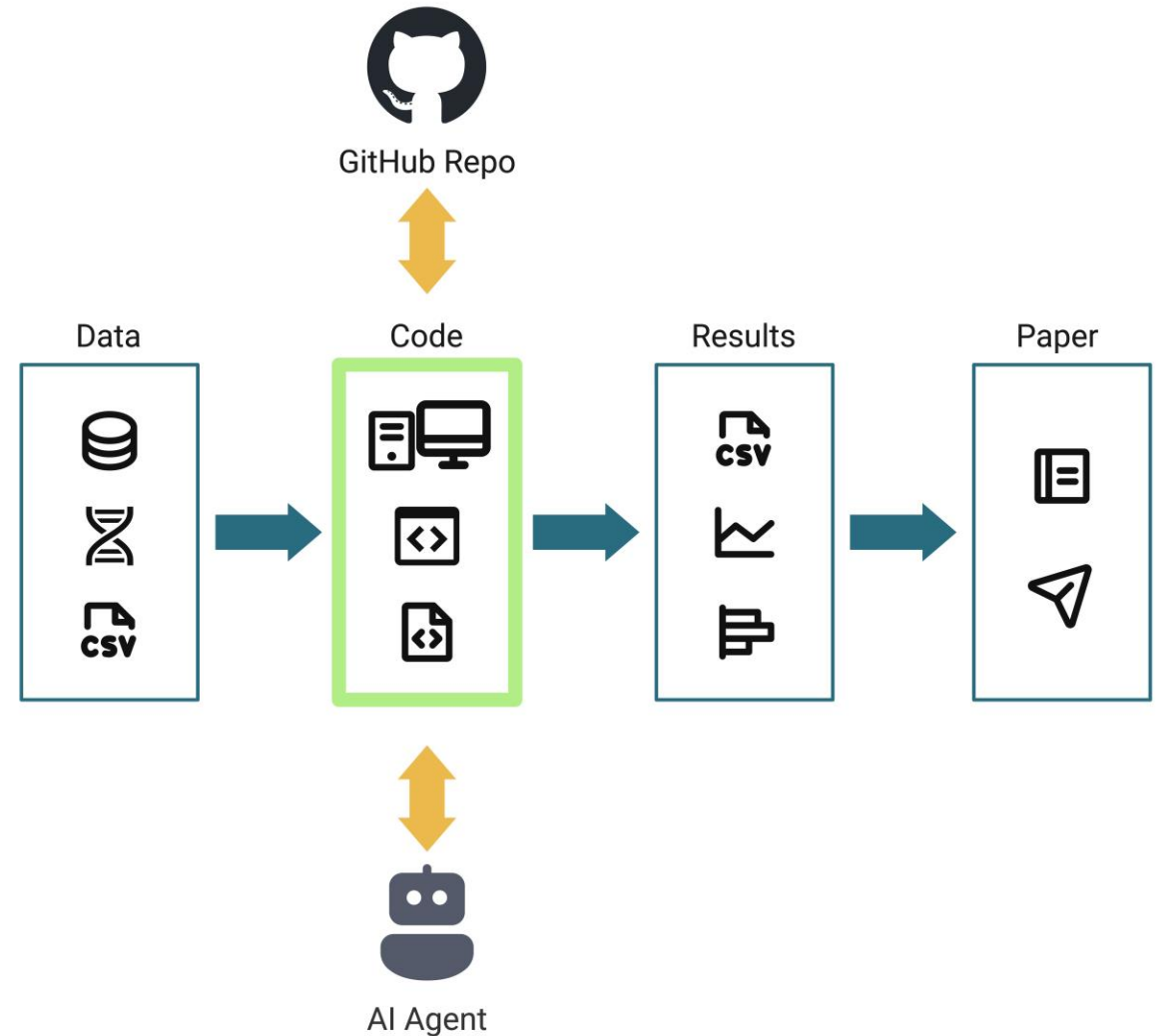
Raw data should be permanently stored somewhere safe (like HPC DME). Use a working copy for your analysis.
Nobody – AI agents or yourself – should edit a raw data file.

Code

Code allows you (or a collaborator or a reader) to reproduce your analysis and plots from raw data. Outputs produced without code are untraceable.

Best practice

Version control using git and sync to GitHub, particularly when working with AI (session 2).



Navigating VS Code

Model selection

Available models

Available models can be viewed by selecting the model list (“Auto”) in Copilot Chat.

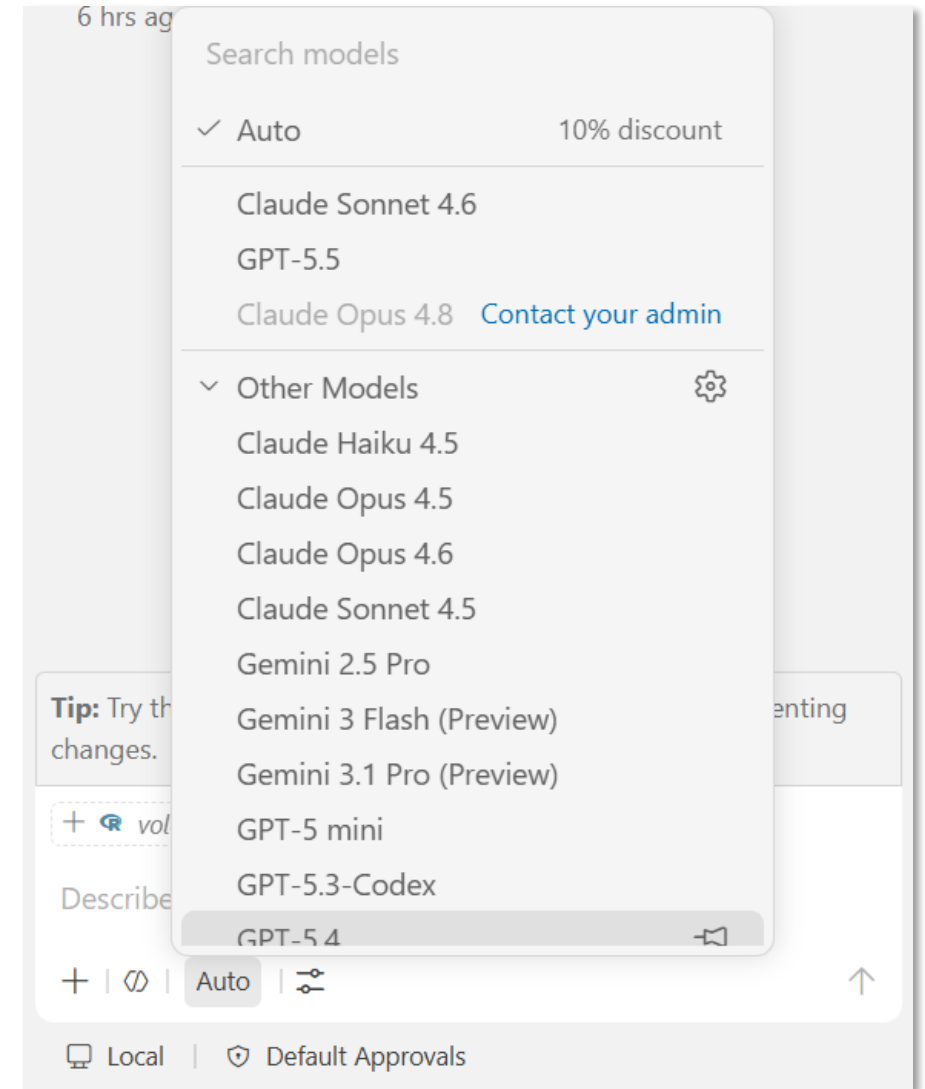
Pricing

Each interaction with Copilot uses tokens / credits.
Different models and actions cost different amounts.

Auto mode

Automatically selects the most efficient available model for each task. Results in a 10% discount.

<https://docs.github.com/en/copilot/concepts/models/auto-model-selection>



Hands-on: Build a volcano plot with Copilot

Input: DESeq2 results table

Output: a volcano plot (.png)

1

Prompt Copilot

Ask Copilot for a short R script to create a volcano plot from your DESeq2 table. Review any proposed actions for security risks, then click Allow.

2

Review the code

Ask questions about any lines that are unclear.

3

Run the script

Ask Copilot to run the script.

4

Review the results

Review the generated plot to see what it produced.

5

Change & rerun

Tweak the code and rerun. If it breaks, ask why and have it help you fix it.

Example dataset

- *PURPL* is a lncRNA and regulator of p53
- Performed RNAseq +/- *PURPL* depletion (sgNTC vs. sg*PURPL* CRISPRi) and +/- p53 induction (Nutlin treatment)
- *PURPL* depletion has widespread transcriptome effects only in the context of p53 induction
- Download Supplementary table 3 (Excel)
- The first 4 tabs contain DESeq2 results (differentially expressed genes between two conditions)
- Tables correspond to volcano plots in Figure 5D
- Optional: Save one table as a tab-delimited text file

► Mol Cell Biol. 2022 Nov 7;42(12):e00289-22. doi: [10.1128/mcb.00289-22](https://doi.org/10.1128/mcb.00289-22) 

Context-Dependent Function of Long Noncoding RNA *PURPL* in Transcriptome Regulation during p53 Activation

[Corrine Corrina R Hartford](#)^a, [Roshan L Shrestha](#)^b, [Lorinc Pongor](#)^c, [Yongmei Zhao](#)^d, [Xiongfong Chen](#)^e, [Caroline Fromont](#)^e, [Ritu Chaudhary](#)^f, [Xiao Ling Li](#)^a, [Katherine R Pasterczyk](#)^a, [Ravi Kumar](#)^a, [Bruna R Muys](#)^a, [Dimitrios Tsitsipatis](#)^g, [Raj Chari](#)^h, [Myriam Gorospe](#)^g, [Mirit I Aladjem](#)^c, [Javed Khan](#)ⁱ, [Munira A Basrai](#)^b, [Ioannis Grammatikakis](#)^{a,✉}, [Ashish Lal](#)^{a,✉}

► Author information ► Article notes ► Copyright and License information

PMCID: PMC9753727 PMID: [36342127](https://pubmed.ncbi.nlm.nih.gov/36342127/)

Supplemental file 3

Table S3. Download [mcb.00289-22-s0003.xlsx](#), XLSX file, 8.1 MB (8.1MB, xlsx)

Copilot cannot replace coding knowledge

Copilot lowers the barrier to coding — it doesn't remove it. If you generate code you can't read, you can't catch when it's wrong. Building a foundation alongside the tool is essential.



BTEP video archive

Recordings of past classes covering R and bioinformatics fundamentals.

<https://bioinformatics.ccr.cancer.gov/btep/btep-video-archive-of-past-classes/>



Introduction to R for novices

An introductory course for those new to R.

<https://bioinformatics.ccr.cancer.gov/btep/courses/introductory-r-for-novices-getting-started-with-r-1>

Additional uses for AI



Research questions

For open-ended research questions, use GitHub's web chat directly — no project setup or credits used, and it runs while you work elsewhere. <https://github.com/copilot>



Understand existing code

Each GB OMICS project has a code base on GitHub. AI can walk you through to learn how the analysis worked.



Reanalyze a finished project

Pull a completed project down from HPC DME and use AI to help you re-run or extend the analysis.

Final thoughts



AI produces plausible-looking output — not understanding

Generative AI uses statistical models to generate what looks like a likely answer to your prompt. Interpreting what that output actually means is your responsibility.



AI accelerates bioinformatics

It makes you faster — just make sure you're facing the direction you want to go. Speed in the wrong direction is still the wrong direction.

Acknowledgements

CCR Collaborative Bioinformatics Resource (CCBR)

Kelly Sovacool-Caruthers

Vishal Koparde

Uma Mudunuri

CCR Genetics Branch

Javed Khan

Paul Meltzer

Natasha Caplen

Ashish Lal

Munira Basrai

Questions?

Reach out to Vineela Gangalapudi or Erica Pehrsson

FUTURE SESSIONS

Session 2: Biowulf and git

- ✓ VS Code on Biowulf to access larger datasets and powerful computational resources
- ✓ Git/GitHub for code version control, reproducibility, and collaboration
- ✓ Custom instructions & skills to steer AI



VS Code on Biowulf via HPC OnDemand



Go to hpcondemand.nih.gov

Open HPC OnDemand in your browser and find VS Code under Interactive Apps.



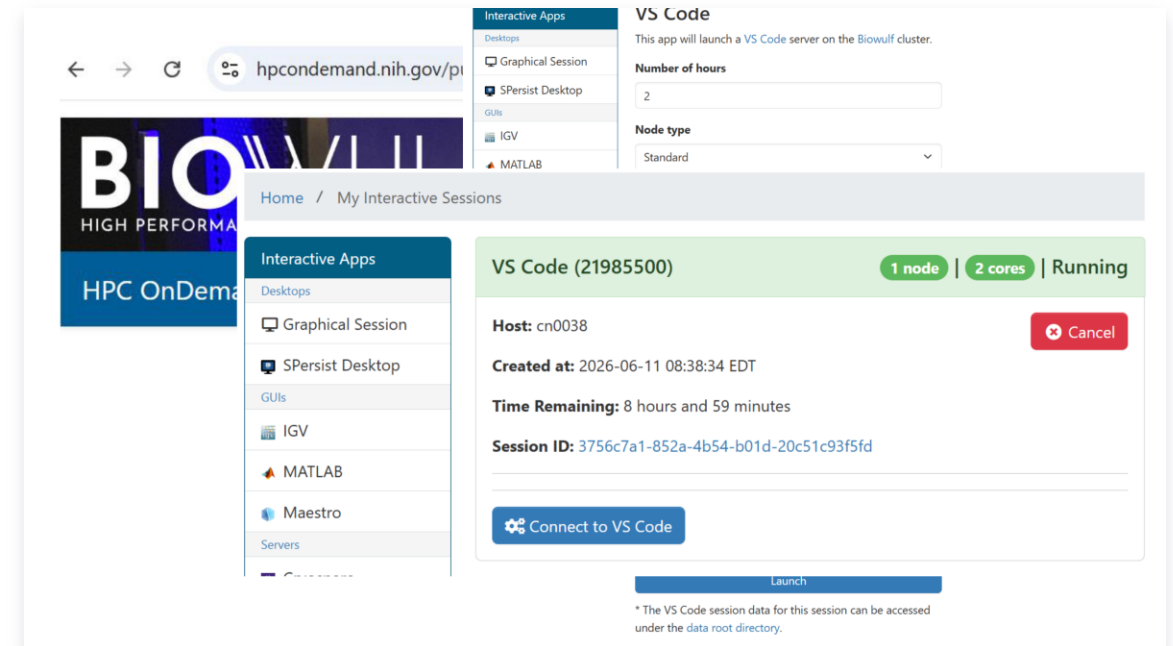
Launch with defaults

Pick the standard node, launch, and wait for the session to start.



Connect to VS Code

Once it's running, click “Connect to VS Code” to open the editor.



HPC OnDemand → VS Code session running on a Biowulf compute node