

Cloud-SPAN Prenomics:

lowering barriers to participation
in big data bioinformatics

by Evelyn Greeves

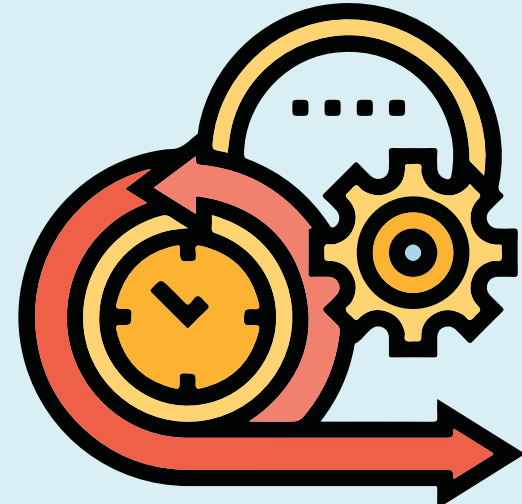
Why learn command line?



improve
reproducibility



reduce human
error



automate repetitive
tasks

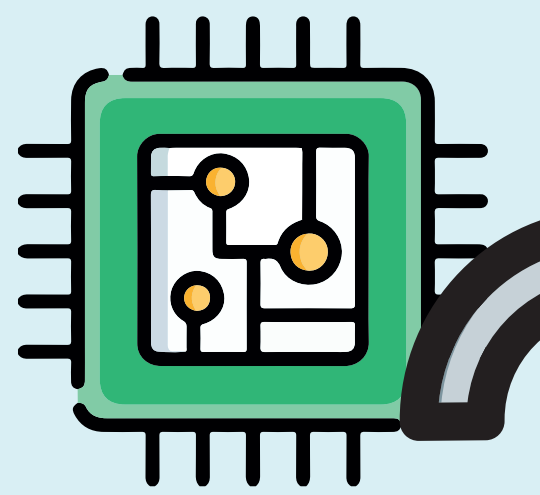


access new tools

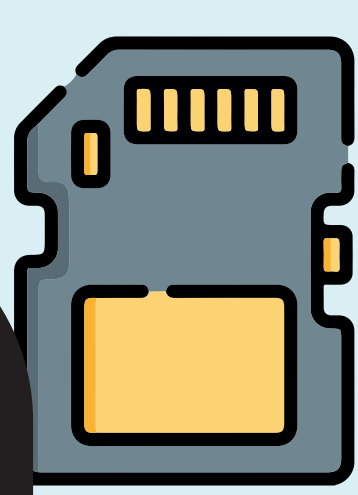
Knowledge of the command line is generally required to access high performance computing resources such as cloud computing.

Why use the cloud?

processing power



memory



disk space



network bandwidth

access more hardware resources



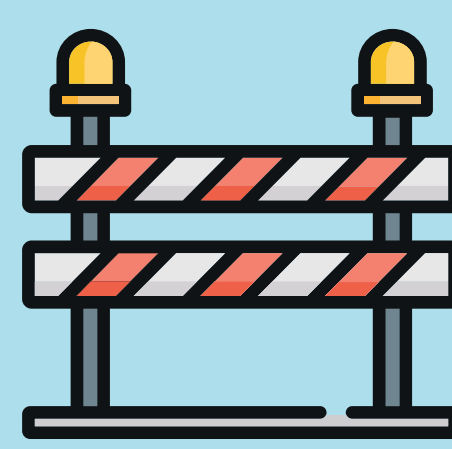
use pre-installed
software



shorten analysis
times



share software/
data containers



overcome barriers
to accessing HPC*

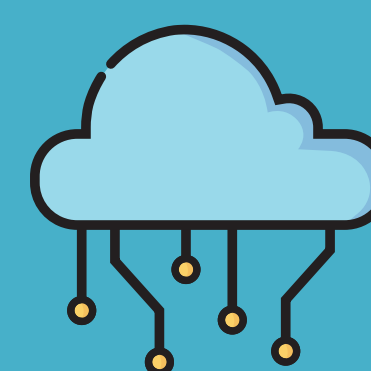
*HPC = high performance computing

What does Prenomics teach?



files & directory
structure

- creating, moving and deleting files using the file explorer
- working directories, absolute paths and relative paths



logging onto
the cloud

- opening the terminal
- changing access permissions for login key file
- logging in using ssh



navigating using
the command line

- using cd, ls and pwd to navigate
- viewing files using cat
- tab completion



working with files
in the command line

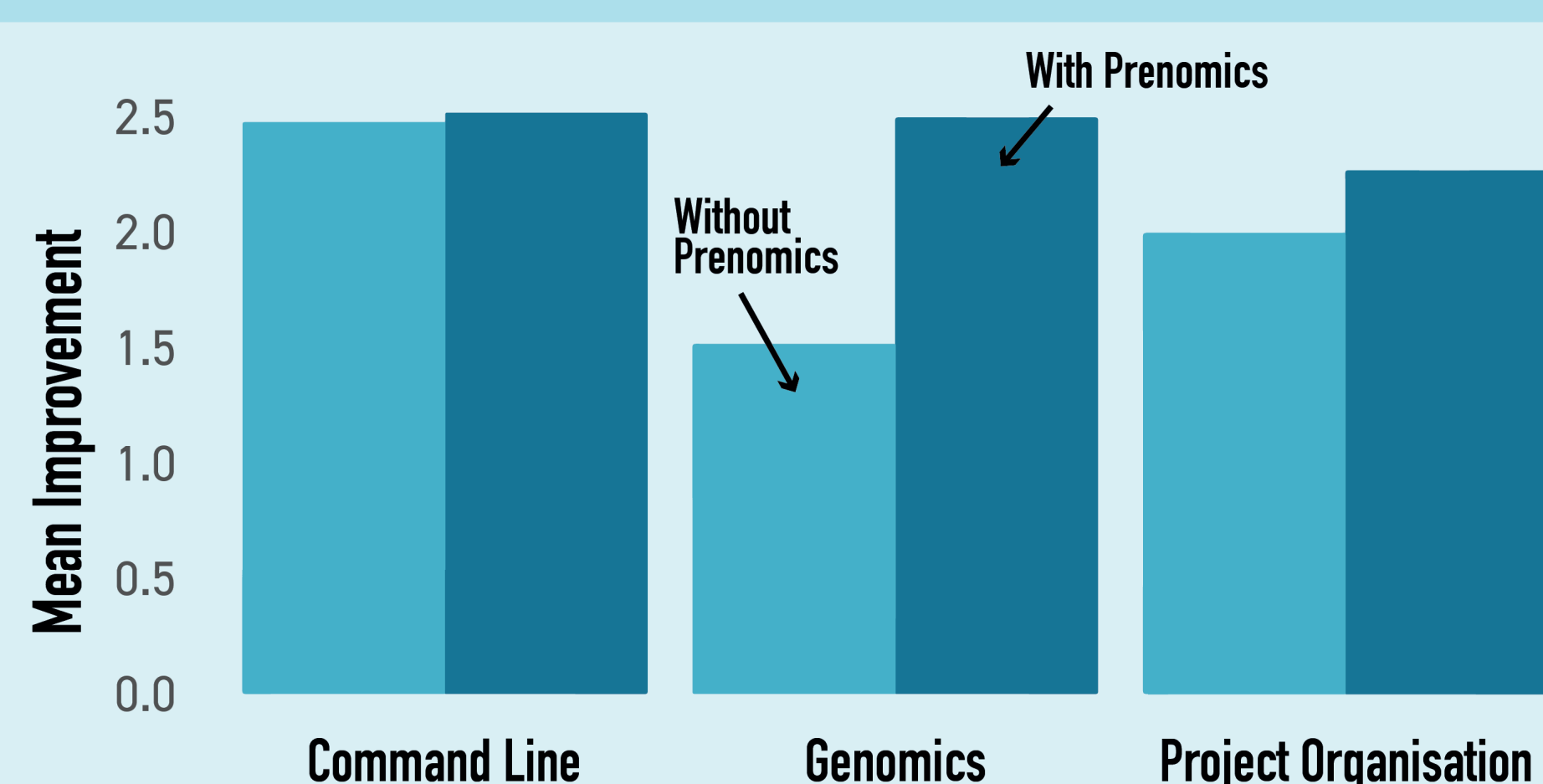
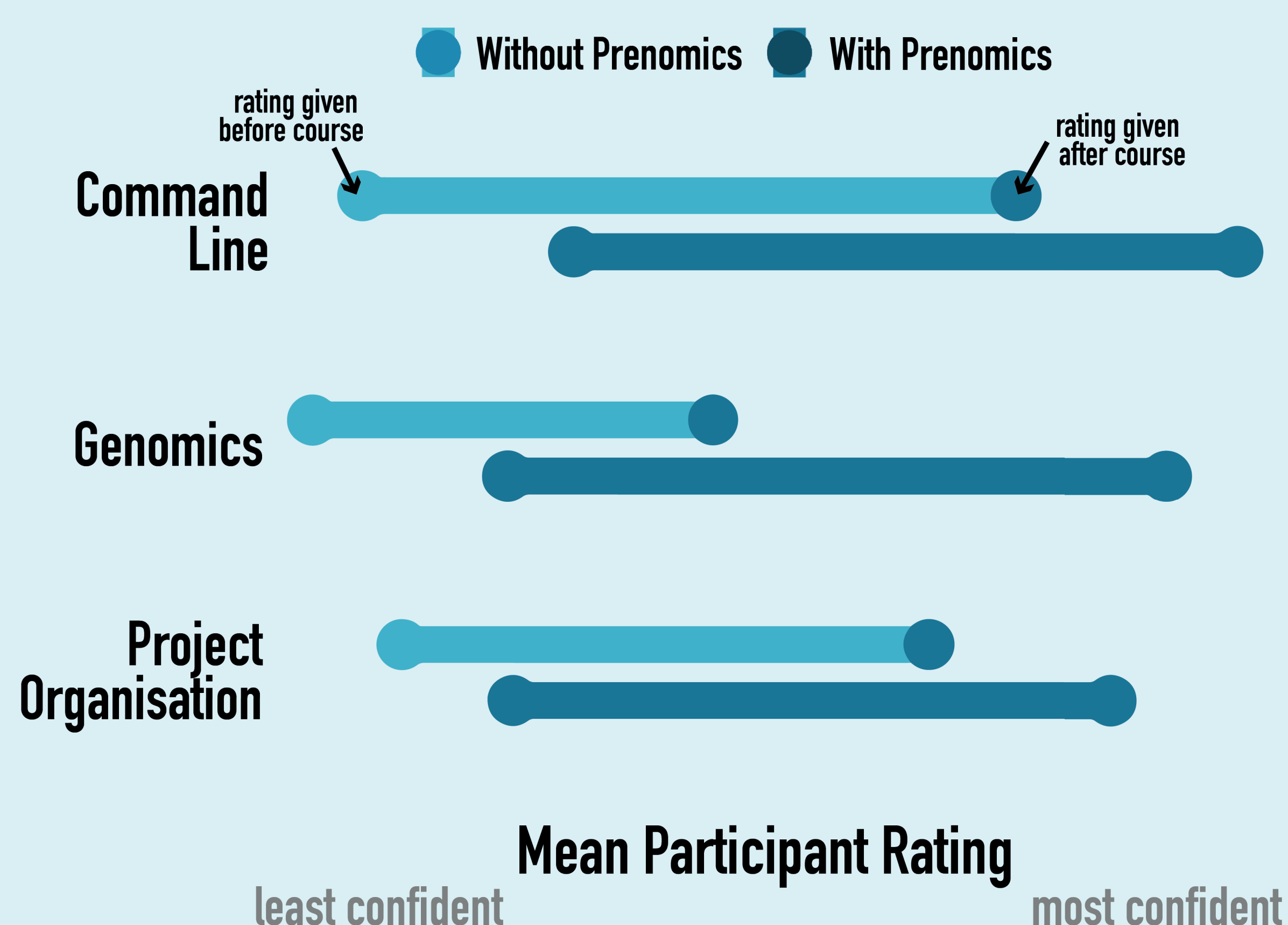
- copying, renaming and moving files
- creating and deleting files
- editing files in nano



searching using
the command line

- using grep to search files
- redirecting output to a file
- redirecting output to a new command

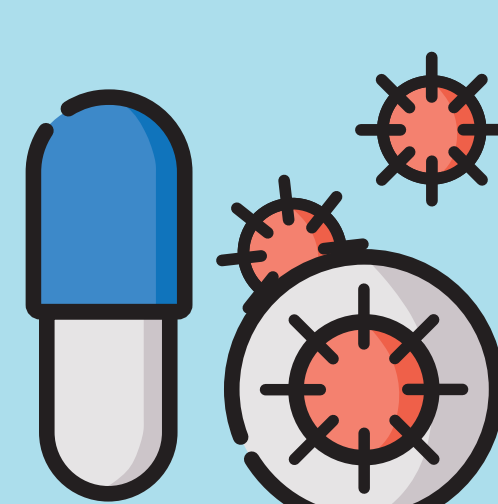
What do the data say?



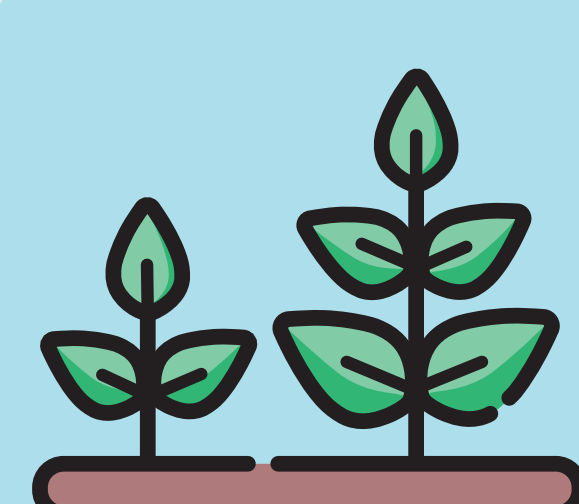
Where is this being applied?*



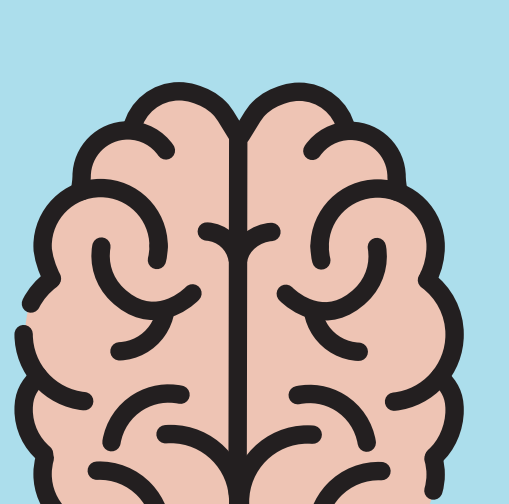
anaerobic
digestion



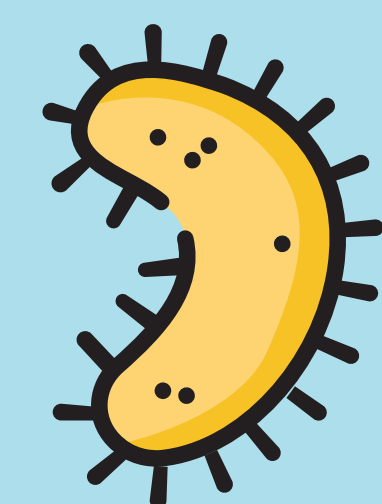
antimicrobial
resistance



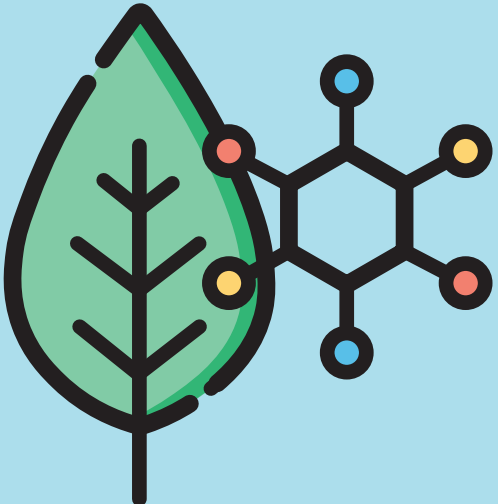
plant evolution



neuroscience



microbial
communities



crop
engineering



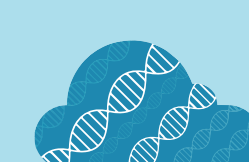
pharmaco-
genomics



biotechnology

*based on backgrounds of course participants

Cloud-SPAN is a collaboration between the University of York and The Software Sustainability Institute, funded by the UKRI Innovation Scholars award. Project Reference: MR/V038680/1



Cloud-SPAN
Cloud-based High Performance Computing
SP: Scalable Pipelines for Genomics



UNIVERSITY
of York



Software
Sustainability
Institute



UK Research
and Innovation

Cite:

Buenabad-Chavez, J., Cansdale, A., Chong, J., Forrester, S., Greeves, E., & Rand, E. (2022). Cloud-SPAN Prenomics Course (Version 0.1.0) [Computer software]. <https://doi.org/10.5281/zenodo.6354056>



<https://cloud-span.github.io/prenomics00-intro/>



<https://github.com/Cloud-SPAN/prenomics00-intro>