

# nature

---

[nature](#) > [career column](#) > article

CAREER COLUMN | 04 October 2023

# Embracing the command line: my unexpected career in computational biology

**A crash course in bioinformatics put Ming Tommy Tang on a different path.**

[Ming Tommy Tang](#)



Ming Tommy Tang wants to help other scientists make the leap to computational biology. Credit: Jiaming Yu

---

Sitting at my office desk behind a large screen and a Macbook, I am crunching single-cell gene-expression data. But there's no arrow pointer to be seen, because I'm not using my mouse. I am analysing these data using written instructions issued in the languages of computational biology: Bash, R and Python.

With more than ten years of experience analysing DNA and RNA sequencing data, I lead the computational biology team at Immunitas Therapeutics in Waltham, Massachusetts. I share computational tips and tricks in blog posts and on X (formerly Twitter), on which I have more than 25,000 followers.

Fifteen years ago, I was a PhD student in a cancer molecular biology laboratory at the University of Florida in Gainesville, and everything was new. I was excited to learn. I clocked at least 10 hours in the lab each day, and quickly became a pipetting expert. I

published my first first-author paper in 2011 and the second in 2013. I was feeling good about my progress.

Then, one day, my adviser asked me to analyse a data set from the Gene Expression Omnibus, a public repository managed by the US National Center for Biotechnology Information. The data were collected using chromatin immunoprecipitation followed by high-throughput DNA sequencing (ChIP-seq), a genome-scale technique for mapping the binding sites of DNA-binding proteins called transcription factors, as well as regions enriched in modifications of histone proteins. My graduate adviser wanted me to probe one of these data sets to learn where a transcription factor called hypoxia-inducible factor-1 binds to the human genome.

The file was 2 gigabytes. I downloaded it, but with more than five million rows of data it crushed Excel, and I didn't know what else to do. I realized for the first time that however good my hands were in the lab, I lacked the data-analysis skills that are increasingly essential to modern life science.



The command line is required knowledge for computational biologists. Credit: Getty

My introduction to those skills came unexpectedly. A colleague in the bioinformatics department at the University of Florida had developed a tool to predict alternative messenger RNA 'splice' sites in genes, and a member of his thesis committee asked him to validate his predictions experimentally. I offered to help. I designed 20 sets of the DNA primers that flank the predicted junctions where the splicing occurs, amplified the sequences between them, and separated them on a gel. In most cases, the primers amplified the desired sequences, showing that his predictions were correct. He passed his defence.

As a token of appreciation, my friend's graduate adviser asked how he could help me in return. I said I wanted to learn bioinformatics, so he gave me a crash course, demonstrating textual commands to sort terms, identify unique values and manipulate tabular data, among other things. It wasn't much, but it was the first time I had seen someone interacting with the computer in this way, and I was hooked. I decided on a change of plan: I would become a computational biologist.

## Brave new world

To newcomers, the text-based command line – called the terminal – can seem scary and unintuitive relative to the drag-and-drop simplicity of modern graphical user interfaces. But it was important that I learn it. For one thing, the analyses that my adviser wanted could not be conducted any other way. Most bioinformatics tools are written to run at the command line. And when using high-performance computing clusters or working in the cloud you have no choice – these computers have no graphical interface. Plus, these terse commands are incredibly good at text manipulation, and when it comes to bioinformatics, text files are the coin of the realm. By chaining simple commands together using the pipe symbol ('|'), bioinformaticians can wrangle plain text files into the desired format to feed into their workflows.

The command line is baked into Unix-Linux operating systems. Users of macOS can access it through the Terminal application, whereas users of Windows 10 and 11 can install the [Windows Subsystem for Linux](#). (Users of older versions of Windows must manually create a dual-boot system, as I did.)

The command line, I realized, would propel me towards computational biology, but it was a rabbit hole. I started to pile up books on my bookshelf. I spent hours setting up a dual-boot system to load Linux on my Windows machine. And I started reading online tutorials and books to learn the basics.

Two resources proved invaluable. The first is an online course on the Unix shell from The Carpentries, an organization in Oakland, California, that provides [workshops on data](#)

[analysis in science](#). The second is the online book, [The Linux Command Line](#) (2019).

Newcomers can also check out my own ebook, [From Cell Line to Command Line](#) (2022).

Even with those aids, don't be surprised if you run into trouble. Linux commands feature unintuitive syntax with confusing and sometimes inconsistent parameters, and it can take months of practice to become proficient. As one anonymous person quoted in *The Art of Unix Programming* (2003) said, "Unix is user-friendly – it's just choosy about who its friends are." In other words, Unix isn't intuitive, until it is; it just takes practice.

As my learning progressed, I found myself at the keyboard more and using my pipettes less. And, once I nailed the basics, I transitioned to the R and Python programming languages and completed my transformation. I did a postdoc in computational biology at the MD Anderson Cancer Center in Houston, Texas, followed by a non-tenure track position at the Dana-Farber Cancer Institute in Boston, Massachusetts, where I led a computational team to analyse single-cell and clinical trial sequencing data.

Ten years after starting my journey towards the command line, I lead a computational biology team at a drug-development company. It wasn't always easy; I was the only one on my floor learning it back in Florida and had no one to turn to for help. I was lucky to have helpful colleagues during my training in Houston who taught me advanced skills, but I needed to work most things out myself.

Through that experience, I learnt the importance of being open-minded and genuinely curious. I now embrace every challenge with determination and discipline, confident that I have the tools and skills necessary to succeed. I am also dedicated to helping other wet-lab biologists make the same transition that I did. If you'd like to make the leap yourself, check out my [blog](#).

doi: <https://doi.org/10.1038/d41586-023-03120-4>

*This is an article from the Nature Careers Community, a place for Nature readers to share their professional experiences and advice. [Guest posts are encouraged](#).*

---

## Latest on:

[Bioinformatics](#)

Computational biology and bioinformatics

Genomics

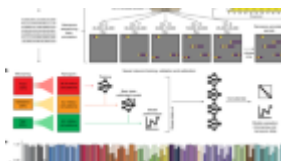


**Methods section too short? Use online protocols to make complex techniques understandable**

CAREER COLUMN |

16 OCT 23

---



**Ultra-fast deep-learned CNS tumour classification during surgery**

ARTICLE | 11 OCT 23

---



**How to spice up your bioinformatics skill set with AI**

CAREER GUIDE | 04 OCT 23

---

Nature (*Nature*) | ISSN 1476-4687 (online) | ISSN 0028-0836 (print)