Computational Biology Upskilling

Tommy Tang, AstraZeneca
Dean Lee, Novartis
June 24, 2025

Mentimeter Survey

Overview

- The need to learn computational skills
- Our learning journeys
- Projects + courses for upskilling
- Design your own upskilling project
- Q&A

The century of biology

*If the 20th century was the century of physics, the 21st century will be **the century of biology**. While combustion, electricity and nuclear power defined scientific advance in the last century, the new biology of genome research—which will provide the complete genetic blueprint of a species, including the human species—will define the next.*

- Craig Venter and Daniel Cohen, November 2004

The century of biology needs computationally equipped biologists

- Biology is becoming increasingly more data-intensive.
- All biologists need at least basic computational skills to handle this data.
- More and more job postings for bench scientist roles also ask for basic computational skills. You don't need to work officially as a computational biologist to benefit from having these skills.
- Biology education in the US has not kept up with this need, still mostly designed for pre-meds.
- So biologists have to figure out how to acquire computational skills on their own.

LinkedIn Conversation, May 14

Darren Nelson · 1st
Founder & CEO | Life Sciences & Technology Talent Partner | Recruitmen...
Book an appointment
1d · Edited · 🌐

Biotech is no longer just wet lab vs. dry lab.

It's wet/dry/data — all at once.

And hiring someone who only checks one box? That's a risk.

At Recruits Lab, we specialize in hybrid talent:
🔹 Bench experience
🔹 Computational fluency
🔹 Translational thinking

Because your next hire shouldn't just fit the job — they should redefine it.

#BiotechHiring #LifeSciences #Interdisciplinary #RecruitsLab

Renee Ng and 113 others   22 comments · 1 repost

Dean Lee · You   (edited) 22h
Figure One Lab: A Gateway Computational Biology Experience...

Hi Darren Nelson I have an idea I'm hoping you can verify. I have anecdotally noticed that more bench scientist positions now ask for Python/R data analysis skills in the "preferred" section. Do you have data that confirms this? Would you say that this reflects a larger trend for the entire industry? Asking for my own curiosity, since I write about biologists learning computational skills. Asking also because you seem to have built a business on this theme.

Like · 🔥 15   Reply · 3 replies   10,417 impressions

Darren Nelson [Author]   11h
Founder & CEO | Life Sciences & Technology Talent Partner | Recr...

Great observation, Dean — and yes, you're absolutely on the mark.

While I don't have a formal dataset to cite (yet!), we're seeing this trend consistently across the roles we fill at Recruits Lab. Bench roles increasingly list Python or R as "preferred," and in many fast-paced biotech teams, they're quietly becoming essential. It's part of a broader shift: data literacy is no longer just for bioinformaticians — it's becoming a core competency for modern scientists.

You've built Figure One Lab around this exact need, and it's clearly resonating. The ability to bridge experimental and computational biology is quickly becoming a career accelerator — and a hiring priority.

Thanks for sparking the convo. Would love to collaborate sometime!

Like · 🔥 1   Reply

Dean Lee · You   11h
Figure One Lab: A Gateway Computational Biology Experience... | 1...

Darren Nelson Good to know I'm not just imagining things!

Like | Reply | 79 impressions

Darren Nelson [Author]   11h
Founder & CEO | Life Sciences & Technology Talent Partner | Recr...

Absolutely, Dean — your instincts are spot on.

We're watching this evolution in real time. In fact, some of our clients have started using coding fluency as a tie-breaker between otherwise equally qualified bench scientists. It's a subtle but growing signal — and I think you're capturing the shift perfectly with Figure One Lab.

Like · 🔥 1   Reply

Computational skills are desired even for bench scientist roles

- https://www.linkedin.com/jobs/view/4215031657
- Scientist, Morphological Profiling/High-Content Imaging (Calico Life Sciences)
  - Advanced tissue culture models including any/all of the following: iPSC-derived disease models, spheroid/organoid models, co-culture systems
  - Plate-based phenotypic screening
  - CRISPR gene editing methods
  - Unbiased profiling technologies such as RNA-seq, proteomics or metabolomics
  - Common lab automation, such as liquid handling
  - **Programming in Python and the fundamentals of machine learning concepts**

# Computational skills are desired even for bench scientist roles

- https://www.linkedin.com/jobs/view/4220617345
- Scientist, Hybrid Computational/Synthetic Biology (Manifold Bio)
  - Design, execute, interpret and iterate on novel library-based experiments
  - **Apply statistical methods and machine learning to NGS data to identify novel variants**
  - Provide deepful insightful analyses of team's NGS experiments and advise on next steps
  - Develop novel protein engineering platform technologies

Computational skills are desired even for bench scientist roles

- https://job-boards.greenhouse.io/digitalbiology/jobs/4668330007
- Senior Scientist, Tumor Model Development (Digital Biology)
    - Develop advanced *in vitro* tumor models (e.g., co-cultures, 3D spheroids) to evaluate biologic therapeutics in an immuno-oncology context.
    - Lead the design, development, and validation of *in vitro* assays (e.g., cell and organoid-based functional assays, cytokine release assays) and *in vivo* models (e.g., xenograft and humanized mouse models) to assess therapeutic antibody efficacy and safety.
    - **Background in computational biology or experience collaborating closely with data scientists.**

Dean's stop-and-go learning journey
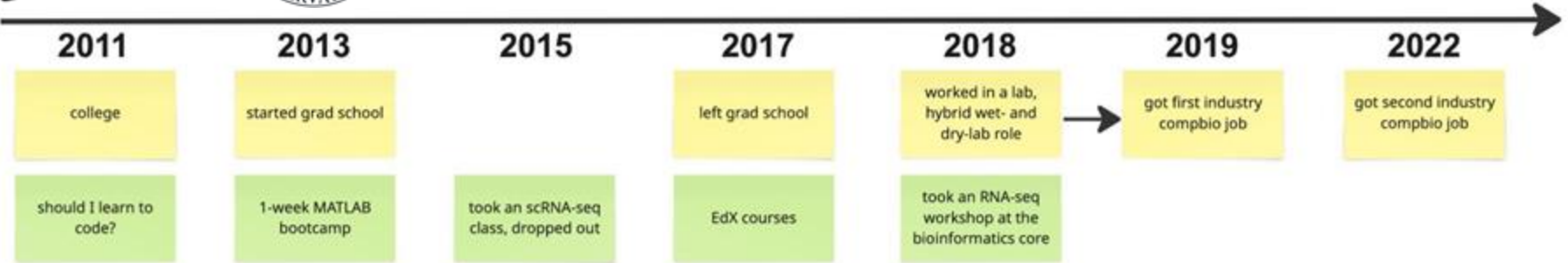
Ming 'Tommy' Tang, PhD

Boston, MA

tangming2005@gmail.com

Director of Bioinformatics

Lead the Bioinformatics effort for
Cancer Immunologic Data Commons

2021- 2024.08.
Director of computational biology

Top 75 Bioinformatics
Blogs by feedspot.com

2020 – 2021
Lead Scientist

Instructor of
Data Carpentries

2018-2020
Senior bioinformatician

Over 40 publications, > 5000
citations

Outstanding
International student

2014-2018
Postdoc & Research Scientist

Blog: https://divingintogeneticsandgenomics.com
30k views per month

2014
**PhD. Genetics & Genomics
(started to learn coding in 2012)**

X @tangming2005 ~39K followers
Linkedin ~40K followers
Youtube: chatomics ~9700

2008 **BS Biotechnology**

Biologists can spend **years** struggling to acquire computational skills.

What if **months** is all it takes to see tangible progress?

Projects + courses to drive learning

Learn by Projects

- Projects is what professional computational biologists do. Every single day.
- Learning to frame biological questions as computational analyses is a valuable skill. It won't go out of style.
- Projects help us put biology, programming, statistics, machine learning, and data storytelling together into one coherent product. That product is still the primary currency in any biotech/pharma context.
- Designing and completing projects appropriate to your level is one of the best ways to upskill.

Learn by Courses

Do courses along the way, while you are working on projects.

Statistics :

- [Modern stats for modern biology](#)
- [Data Analysis for the Life Sciences](#)
- [How I Would Learn Bioinformatics From Scratch 12 Years Later: A Roadmap](#)

Programming:

- R
- Python
- Julia

Tips for Designing An Upskilling Project

- Start with data closest to you, within your biological niche.
- Start with a common data modality, Ex. RNA-seq
- Start with somebody's else's code.
- Start with Python or R, no need to learn both at once.
- ChatGPT is your programming tutor.
- Learn as you complete projects, don't get stuck taking courses/tutorials but never striking out on your own.
- To start, you'll do a lot of plotting. This is not intimidating. It's rewarding!
- Be mindful of compute. You can't handle large datasets from just your laptop. Pick manageable datasets; you can still learn everything you need to learn.

Upskilling Project Worksheet

- Question
  - Keep it very simple at first. Ex. Is Gene X in mouse T cells also expressed in human T cells?
- Data
  - TCGA, CCLE, GTEx, etc.
- Method
  - A GitHub repo or tutorial (Tommy has many) to get you the code to start with.
  - Use existing, popular packages. Don't reinvent the wheel.
- Language
  - Based on what you chose for method
- Courses
  - Classes you might take to shore up knowledge gaps? Many employers offer tuition reimbursement that go unused.

Example of An Upskilling Project (Dean)

- **Background**: I was working in a neuroscience lab at Harvard on a project related to the mouse hypothalamus, NOT working on sleep neurons. This was a stretch goal for me.
- **Question:** What are the molecular subtypes of sleep-active neurons in the mouse hypothalamus?
- **Data:** GSE79108, single-cell RNA-seq of sleep-active neurons from mouse hypothalamus (from Chung et al.)
- **Method:** Seurat
- **Language:** R
- **Outcome**: Reproduced published findings and generated additional, biological hypotheses for future validation.
- **Full story here**: https://www.linkedin.com/pulse/how-i-built-compbio-project-my-free-time-land-biotech-dean-lee-ntxse/
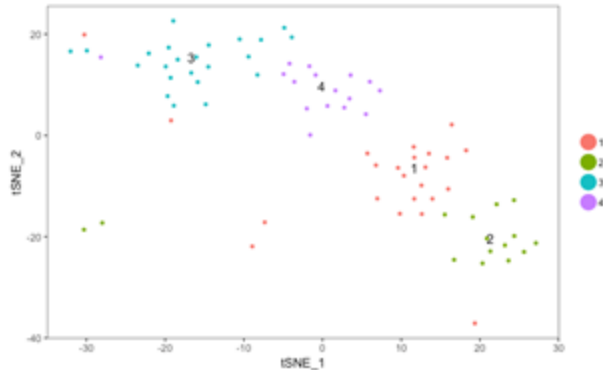
Example of An Upskilling Project (Dean)

Identification of preoptic sleep neurons using retrograde labelling and gene profiling

Shinjae Chung, Franz Weber, Peng Zhong, Chan Lek Tan, Thuc Nghi Nguyen, Kevin T. Beier, Nikolai Hörmann, Wei-Cheng Chang, Zhe Zhang, Johnny Phong Do, Shengin Yao, Michael J. Krashes, Bosiljka Tasic, Ali Cetin, Hongkui Zeng, Zachary A. Knight, Liqun Luo & Yang Dan ✉

ISH images from Allen Brain Institute

Cluster 3: Gpx3, Ngb
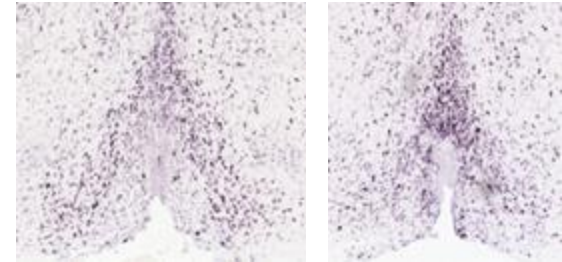
Cluster 4: Chat, Cd44, Slc5a7

The largest group, as previously described by Chung et al., consists of neurons expressing Tac1 and Pdyn. These two markers capture the majority of the sleep-inducing neurons in this study. My further analysis, however, revealed that at least two other groups of neurons, distinct from the Tac1/Pdyn neurons, can be defined from the same dataset.
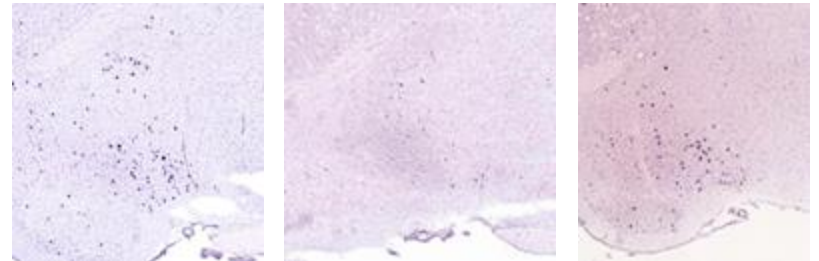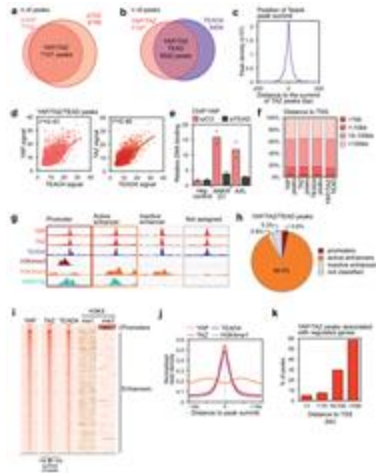
Example of An Upskilling Project (Tommy)

Website:

https://crazyhottommy.github.io/reproduce_genomics_paper_figures/

Github repo:

https://github.com/crazyhottommy/reproduce_genomics_paper_figures

Tip #1 Get on social media

- Get on social media: Twitter/X, Mastodon, Bluesky, LinkedIn
- Follow people of the same interest; bioinformatics papers, AI. (I got to know the most recent AI advances - Replit, Cursor, Lovable, Manus, Owl, etc. - from social media)

I started using Twitter after reading Stephen Turner's blog: How to stay current on bioinformatics: https://www.r-bloggers.com/2017/02/staying-current-in-bioinformatics-genomics-2017-edition/

Tip #2 Write a blog

## I web, therefore I am

Yihui Xie

Why a blog?

- a blog post is much better than a statement "good at R or Bayesian stats" on your CV
- "spend 30 minutes each day in 5 years building a website" vs "20 hours to write a CV in the last semester"
- there are many things that are more suitable for web pages (see my blog for example)

Credit: Yihui Xie

https://slides.yihui.org/2012-stats-web.html

Build a website so others can find you

https://tommytang.bio.link/
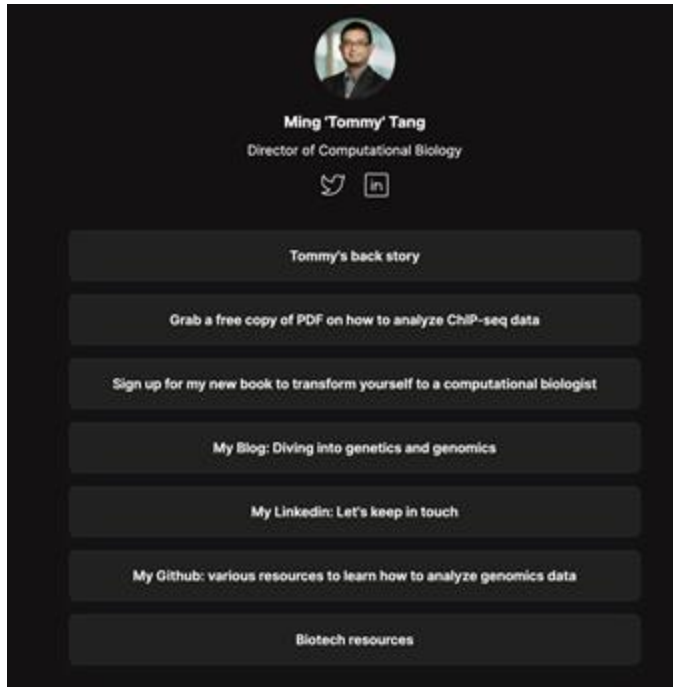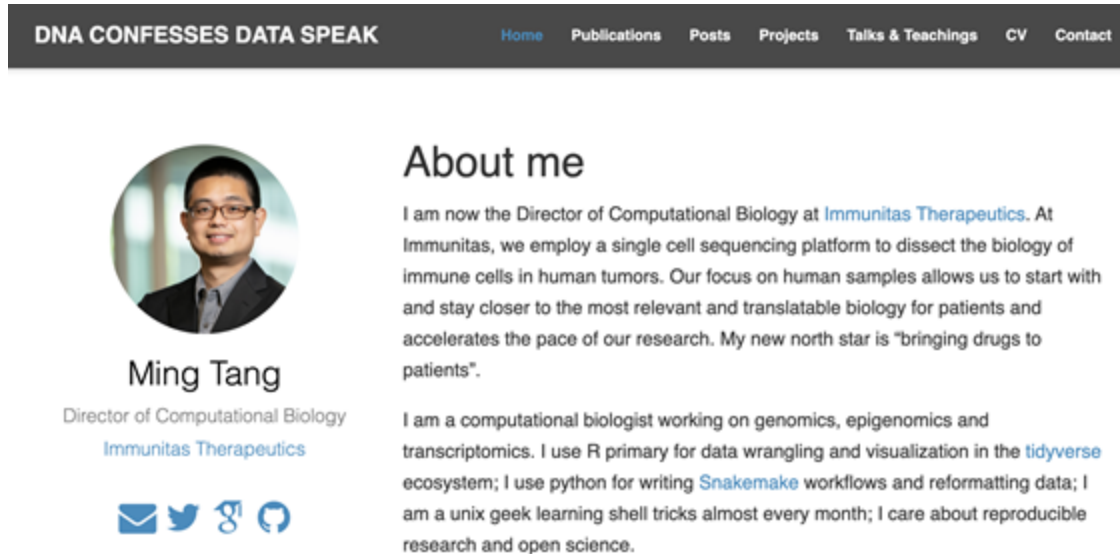
divingintogeneticsandgenomics.com

Start now

- If you do not have a website yet.

- The best time to start one is 10 years ago, the second best time is now.

- Take a weekend to set it up. Be visible: Blogdown or Quarto.

Tip #3 How to connect with people? On social media or in real life

One core thing to remember: Always give value on the table. Be a giver not a taker.

- Complement
- Ask questions, be genuinely curious
- Offer help
- Ask people to meet online

Do not be transactional. Dig the well before you get thirsty.

Follow Us on LinkedIn for More

[Dean's LinkedIn](#)

[Tommy's LinkedIn](#)

Q&A