Data science training at the University of Colorado

Lawrence Hunter, Ph.D.
Director, Computational Bioscience Program
University of Colorado School of Medicine

Larry.Hunter@ucdenver.edu
http://compbio.ucdenver.edu/Hunter
@ProfLHunter

PhD Program

- Focus: Creating developers of novel methods
- Core curriculum:
    - Shared biomedical research "core" course
    - Introduction to Biomedical Informatics
    - Advanced Biomedical Informatics
    - Statistics requirement (various ways to meet)
    - Ethics for Biomedical Informaticians
- Our students matriculate with strong CS (usually MS or industrial experience)

How to stay up to date?

- All core courses (except statistics) are updated annually.

- Advanced biomedical informatics course reflects faculty research interests

- Key pillar of our training program mission statement is that students are trained to become self-directed and life-long learners.
  - Woven into curriculum and training experiences throughout graduate program

Student diversity

- Despite very diverse backgrounds (majority female, significant African-American, Hispanic and Native American enrollment), all of our students matriculate with substantial computational and mathematical backgrounds

- Preparatory work available in our non-degree summerSTTP program (several current students are graduates).

Tools and Techniques

- Our focus is on hands-on research training.
  - Students begin learning about design of research projects on the first day of first semester
  - "Intro" course work is to write an R03-like proposal, do peer review, rewrite, and present orally and in writing.
  - "Advanced" course work is similar, but requires executing proposed project and reporting, orally and in writing, on the results of the project
  - Rotations begin second semester.

Tools and Techniques (2)

- Strong focus on scientific communication:
  - Making a claim and supporting it with evidence
  - How to argue for significance of a claim
  - How to contest a claim and/or evidence
- Lots of practice in written and oral presentations in various genres:
  - Research plans and critiques (grants and reviews)
  - Research results and critiques (papers, manuscript reviewing, oral presentations, asking questions)

Tools and Techniques (3)

- Ethics training a central program component
  - Social context in which research takes place:
    - Who pays for what kind of research, and why?
    - How biomedical informatics fits into society, including an international perspective
    - What are the broader impacts of informatics research?
  - Technical solutions to ethical problems
    - Protected Health Information, and why it matters
    - Privacy, security and encryption, with examples
  - Student selected topics

Postdoctoral training

- More flexible about backgrounds of trainees, but all have decent computational background

- Individual training plans, carefully monitored

- Sometimes involve distance learning for computational skills (including Boulder).

- Coordination with other postdoctoral training programs on campus (D2V: Big Data to Patient Value, Genomic Cardiology fellowship)

Challenges

- Recruiting computer scientists to join our program (often for a personal reason)
- Training skilled bioinformatics analysts
  - Some of our faculty help, but it's not a responsibility of our training program
  - Distributed across different programs (statistics, biochem & genomics, personalized medicine)
- Finances (we could admit and train many more good students, but are limited financially)