Perspectives on the Past, Present, and Future of Bioinformatics Curriculum Design

Russell Schwartz

Computational Biology Department and
Department of Biological Sciences

Carnegie Mellon University

iSCB INTERNATIONAL SOCIETY FOR COMPUTATIONAL BIOLOGY

Carnegie Mellon

My Perspectives on Bioinformatics Education

- Student
  - Learning to be a computational biologist while doing degrees in computer science (because there were no degrees in computational biology)
- Course instructor
  - Developing various courses in computational biology, genomics, data science, and biological modeling aimed at experimental and computational biologists
- Curriculum developer
  - Helping design curricula for computational biology BS, MS, and PhD degrees.
- Program administrator
  - Directing education programs for MS and PhD in Computational Biology and MD/PhD.
- Department head
  - Running a Computational Biology Department offering programs at the pre-college, BS, MS, and PhD levels
- ISCB Education Committee member and COSI co-chair
  - Being part of a community dedicated to improving bioinformatics education internationally

Computational Biology Training at Carnegie Mellon

Degree Programs for Computational Biology Specialists

BS in Computational Biology

4-year degree providing solid foundation in biology, computer science, and the intersection. Students typically go on to graduate study (PhD or MD).

MS in Computational Biology

2-year degree to provide graduate level training in computation and its use in biology. Students typically go on to industry (pharma, biotech, software, etc.) or PhD programs.

PhD in Computational Biology (joint with University of Pittsburgh)

Advanced training for computational biology researchers. Most students go on to academia or research positions in biomedical industry.

Courses for Experimental Biologists

Introduction to Computational Biology (aimed at biology undergraduates)

1-semester sophomore core class to give skills that will help them in more advanced coursework, research projects, and higher study.

Data Science for Biologists (aimed at biology PhD students)

One-semester first year graduate course, designed to prepare new experimental biology students for practical data analysis problems they will encounter in experimental lab work.

Training Computational Biologists: 4-Year BS

Math/Stats Core (5 classes) Calculus, differential equations, concepts of math, general math elective, stats elective	**General Science (4 classes)** General chemistry I, general chemistry II, organic chemistry, physics
Biological Sciences (8 classes) Modern biology, biochemistry I, cell biology, genetics, molecular biology lab, colloquium, seminar, advanced bio elective	**Computer Science (6 classes)** Imperative computing, functional programming, data structures/algorithms, theory, algorithm design/analysis, advanced CS elective
Computational Biology (3-4 classes) Computational genomics, Biological modeling, 1-2 comp bio electives	**General education (14-15 classes)** Computing, communications, 7 humanities electives, 5-6 free electives

Training Computational Biologists: 2-Year MS

Fall Year 1
Programming, Algorithms & Data Structures, Essential Math & Statistics, Applied Cell & Molecular Biology, Professional Issues

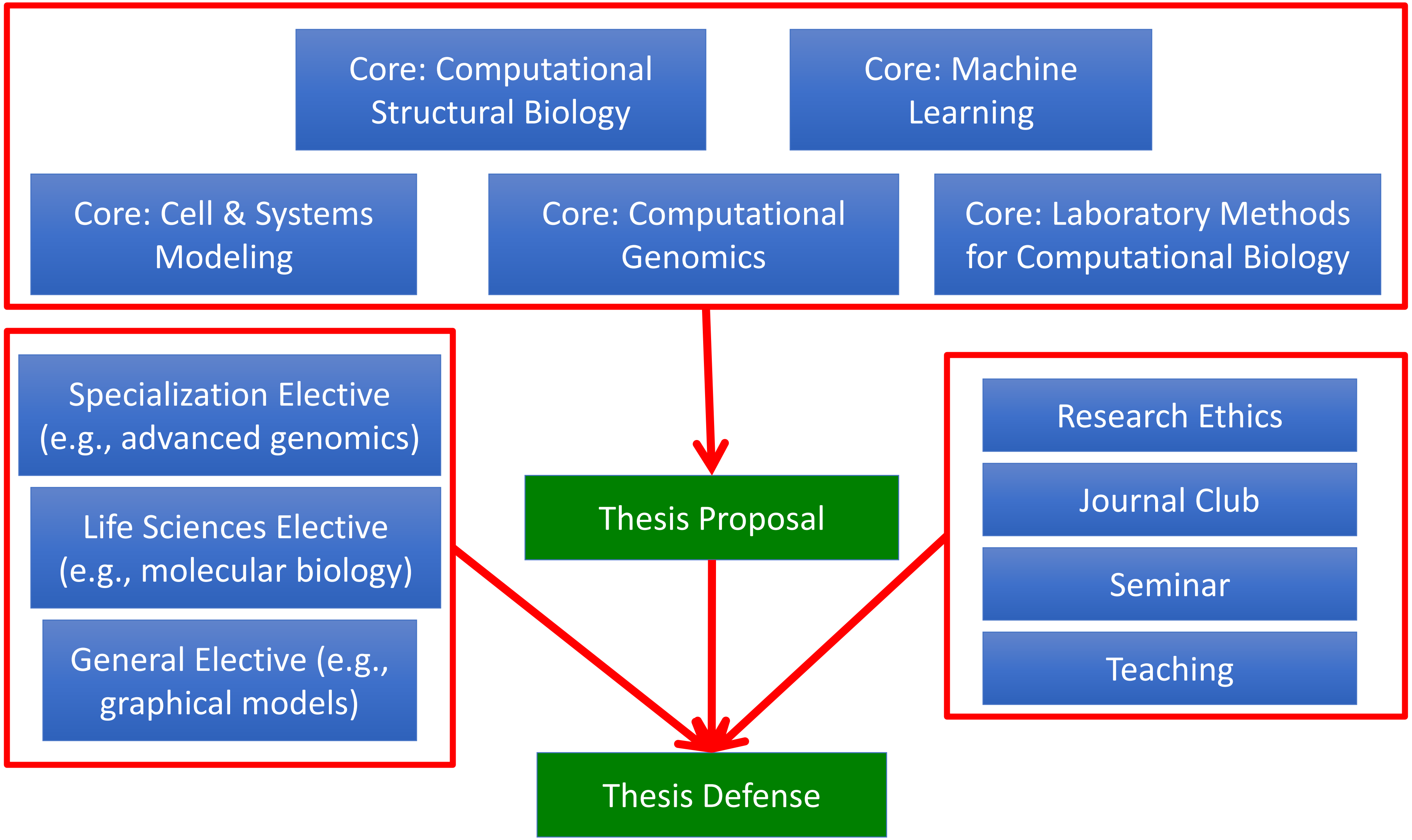
Spring Year 1
Machine learning, Quantitative Genetics, Seminar, Biology or Computational Biology Elective

Summer
Industry internship or on-campus research

Fall Year 2
Biological Modeling & Simulation, 2 x Biology or Computational Biology Elective

Spring Year 2
Automation of Science, Genomics, Biology or Computational Biology Elective

Training Computational Biologists: PhD (typically ~5 years)



Computational Biology Training at Carnegie Mellon

Degree Programs for Computational Biology Specialists

BS in Computational Biology

4-year degree providing solid foundation in biology, computer science, and the intersection. Students typically go on to graduate study (PhD or MD).

MS in Computational Biology

2-year degree to provide graduate level training in computation and its use in biology. Students typically go on to industry (pharma, biotech, software, etc.) or PhD programs.

PhD in Computational Biology (joint with University of Pittsburgh)

Advanced training for computational biology researchers. Most students go on to academia or research positions in biomedical industry.

Courses for Experimental Biologists

Introduction to Computational Biology (aimed at biology undergraduates)

1-semester sophomore core class to give skills that will help them in more advanced coursework, research projects, and higher study.

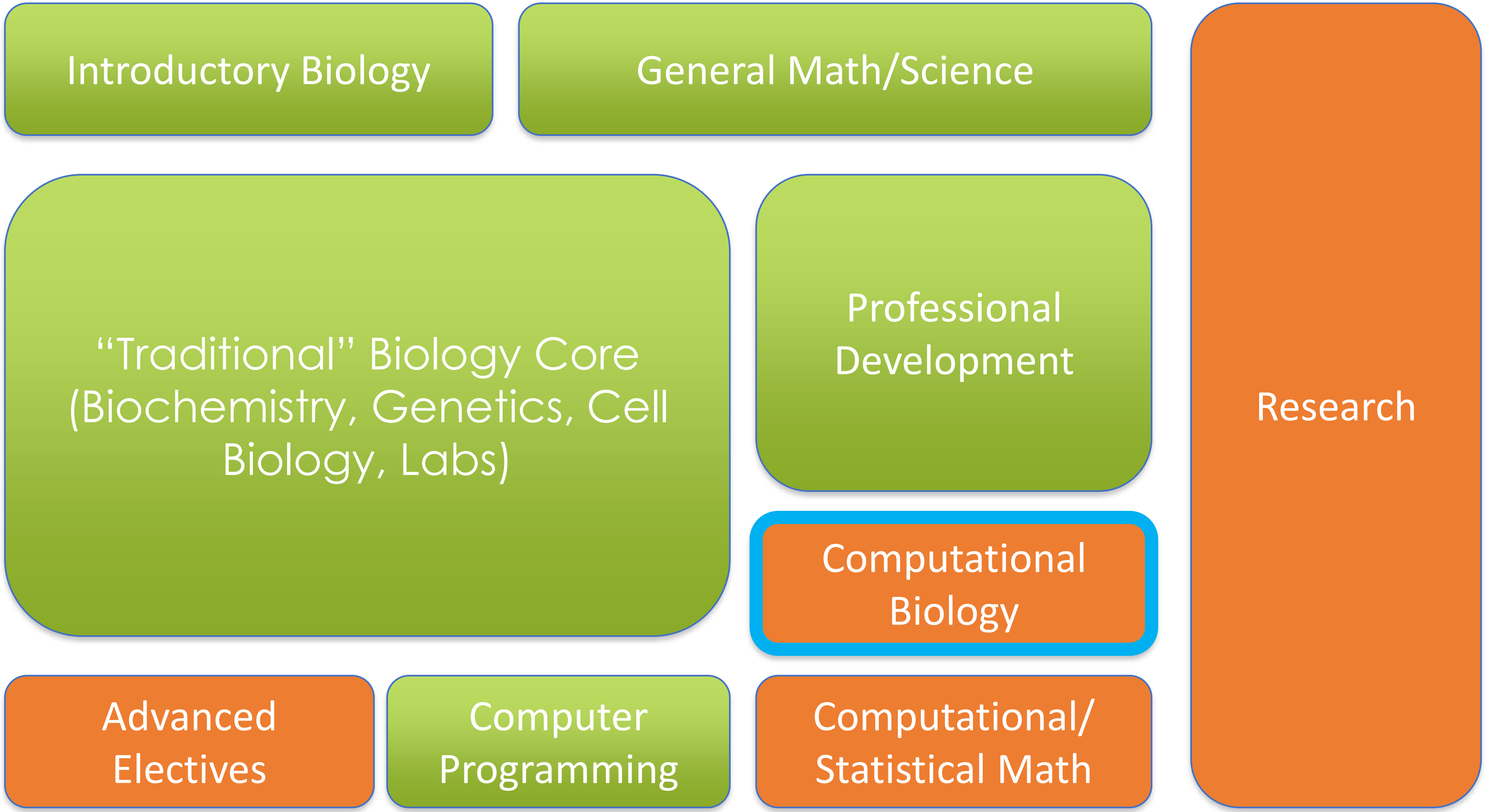
Data Science for Biologists (aimed at biology PhD students)

1-semester first year graduate course, designed to prepare new experimental biology students for practical data analysis problems they will encounter in experimental lab work.

Fitting Computational Biology in an Undergraduate Biology Curriculum: Twenty Years Ago

Required

Elective



Introduction to Computational Biology: Official Learning Objectives

1. Learn major biological data types, the methods by which they are produced, and their uses.
2. Learn to critically assess the reliability of biological data sources.
3. Learn essential concepts of statistics and algorithms needed to productively use database search, analysis, and inference tools and interpret their results.
4. Learn to synthesize results from different data sources and select sources appropriate to a given problem.
5. Learn about of the major repositories of biological data and the tools to access them.
6. Learn to independently research a biological question using online resources.
7. Learn how to pose biological questions through mathematical models and reason about the assumptions and limitations of those models.
8. Learn to simulate the behavior of simple mathematical models.
9. Learn basic tools and concepts of biological image analysis.

Introduction to Computational Biology: Unofficial Learning Objective

Learn what you need to know about computational biology to be the best biologist you can be:

- ▶What computational tools are available to you?
  (algorithms, data sources, ways of thinking)
- ▶What kinds of problems can they help you solve?
  (experimental design, hypothesis testing, making inferences from large data sets)
- ▶What do you need to know about computer science and statistics to use them effectively?
  (abstracting a problem, choosing appropriate methods, interpreting outputs)
- ▶What do you need to know about biology and biotechnology to use them effectively?
  (identifying hypotheses, identifying appropriate data sources, critically evaluating computational results)

Intro to Computational Biology Course Syllabus

Genomics and Molecular Biology

Week 1: Sequences and sequence databases

Week 2: Sequence searching and alignment

Week 3: Protein structure and domains

Week 4: Genome sequencing and annotation

Week 5: Genetic variation and disease

Week 6: Gene and protein expression

Week 7: Networks and pathways (Systems biology)

Week 8: Phylogenetics, Review and Midterm Exam

Modeling and Image Informatics

Week 1: Population dynamics and mathematics of modeling

Week 2: Biochemical kinetics

Week 3: Neuronal models

Week 4: Network models

Week 5: Introduction to statistical machine learning

Week 6: Bioimage analysis I

Week 7: Bioimage analysis II

Week 8: Final exam

Overall Course Design

- **Monday/Wednesday/Friday Lectures:**

Mon: Fundamentals of statistics	Wed: Sequence alignment	Fri: BLAST
-Probability basics	-Scoring functions	-Nucleotide+protein BLAST
-Statistical hypothesis testing	-Significance of alignments	-BLAST statistics
-P-values, E-values, etc.	-Algorithm basics	-Using BLAST in practice

- **Friday Recitation (computer lab):**
  - Class works as a group through a mini-research assignment, using BLAST and other tools and interpreting the results
  - Originally purely web-based tools, later added Python scripting and basic programming
  - Example Topic: Identifying likely sources of a food contaminant

- **Homework:**
  - Mini-research project combining conceptual, mathematical, and hands-on material
  - Example Topic: Inferring origins of a new influenza strain from sequence data

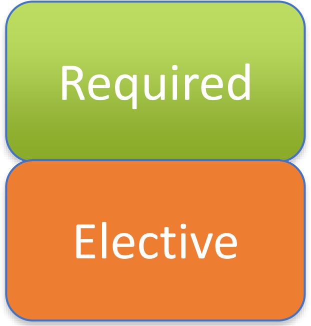
Creating a course is just the first step, however. We also need to …

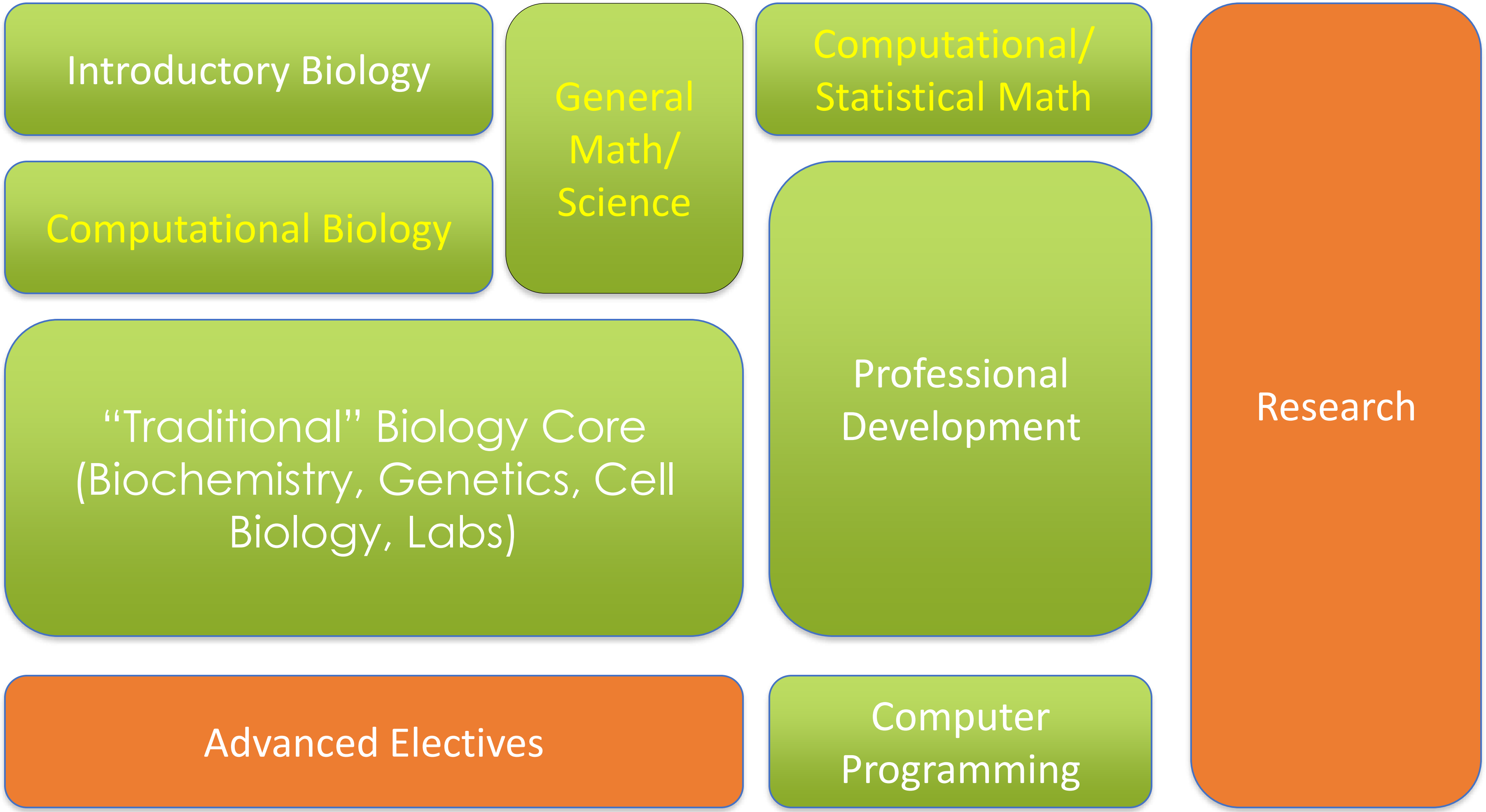
… make room for it in the curriculum

… ensure students can get the necessary prerequisites before reaching it

… ensure students (and subsequent classes) can use what they learn in it

Fitting Computational Biology in an Undergraduate Biology Curriculum: Ten Years Ago

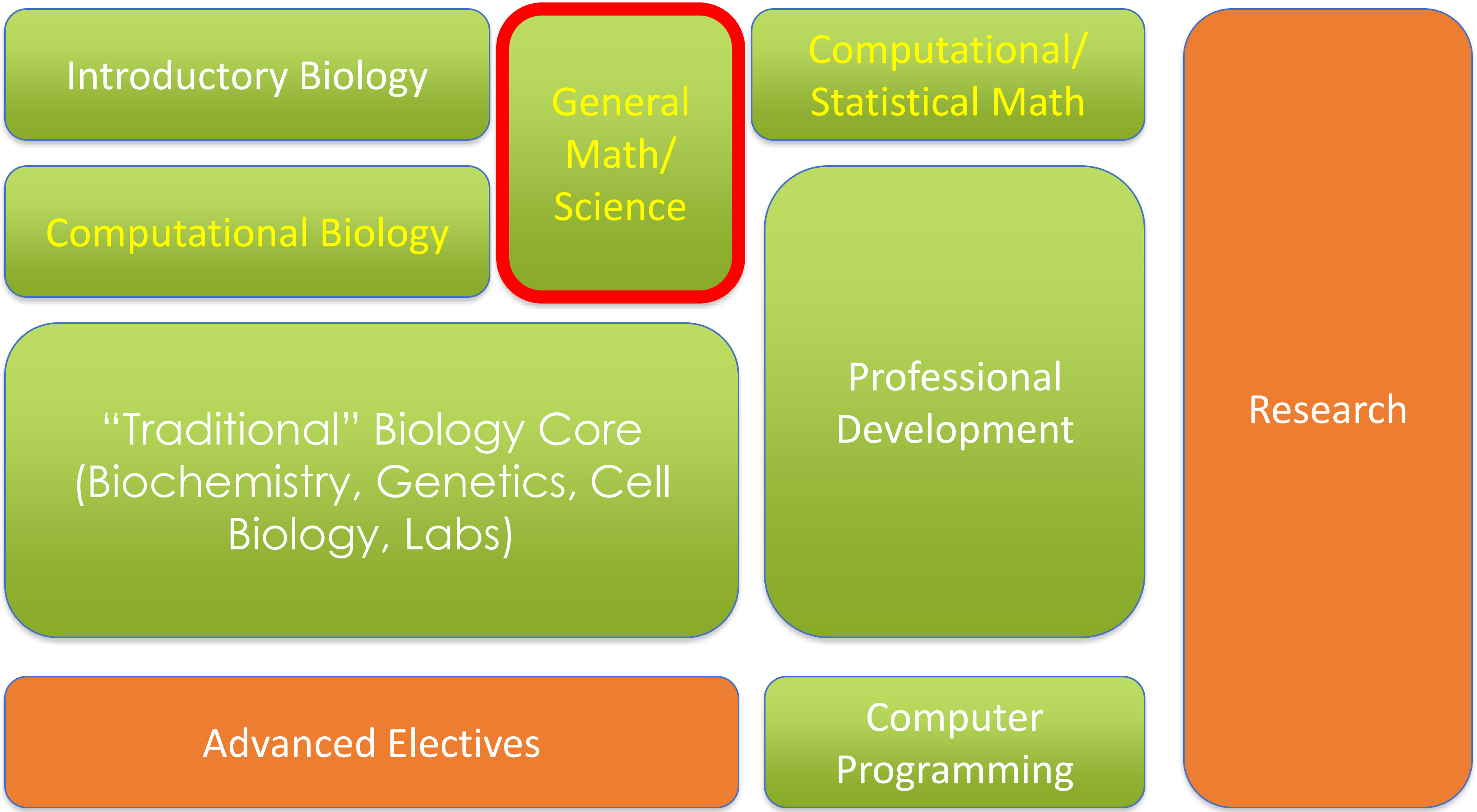




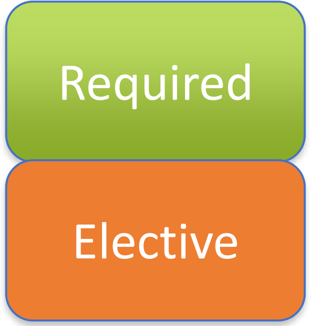
Fitting Computational Biology in an Undergraduate Biology Curriculum: Ten Years Ago

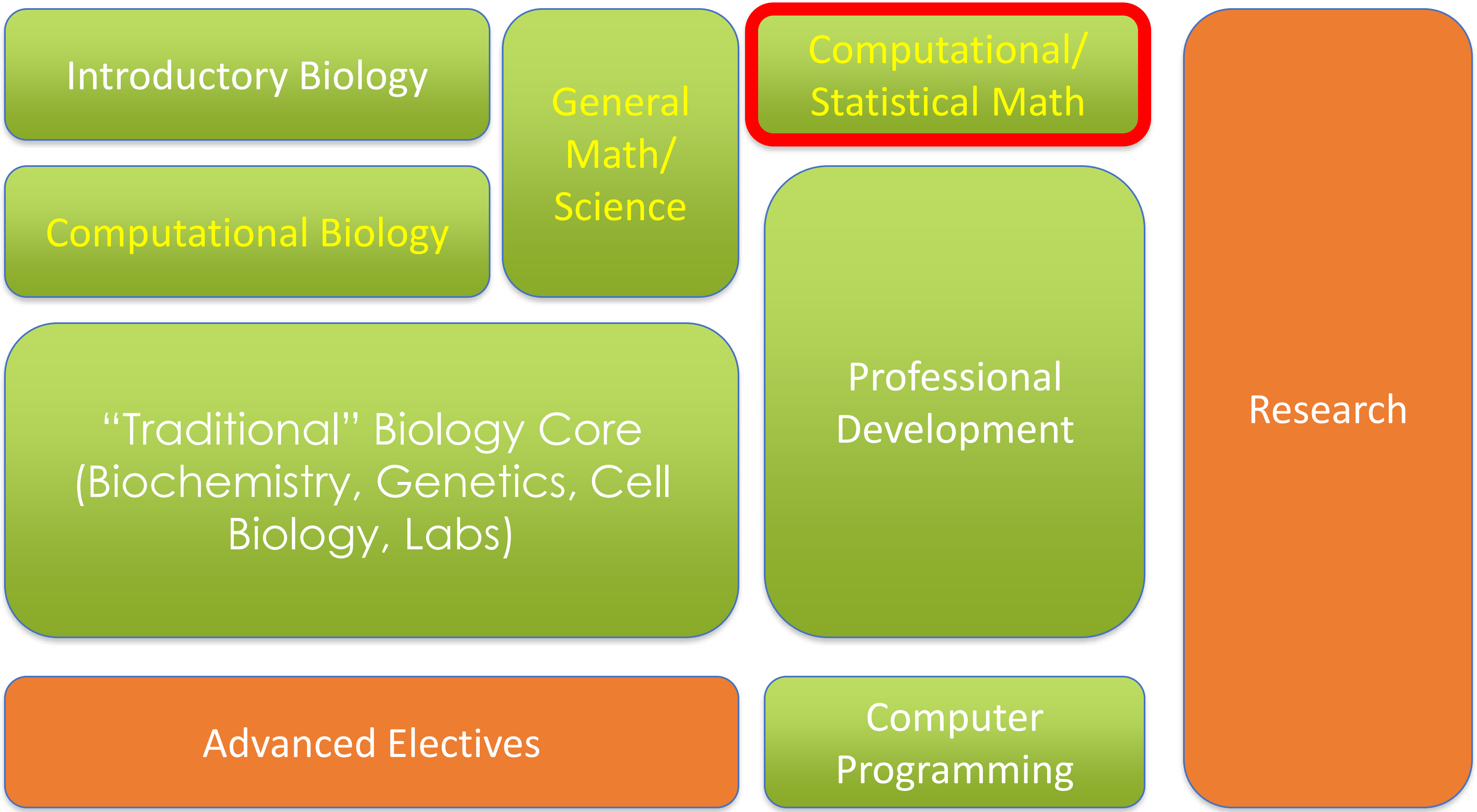
Required

Elective

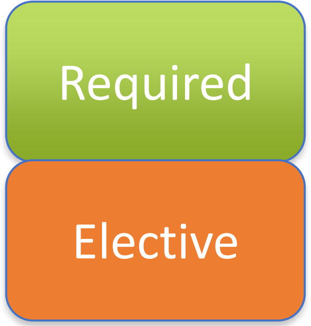


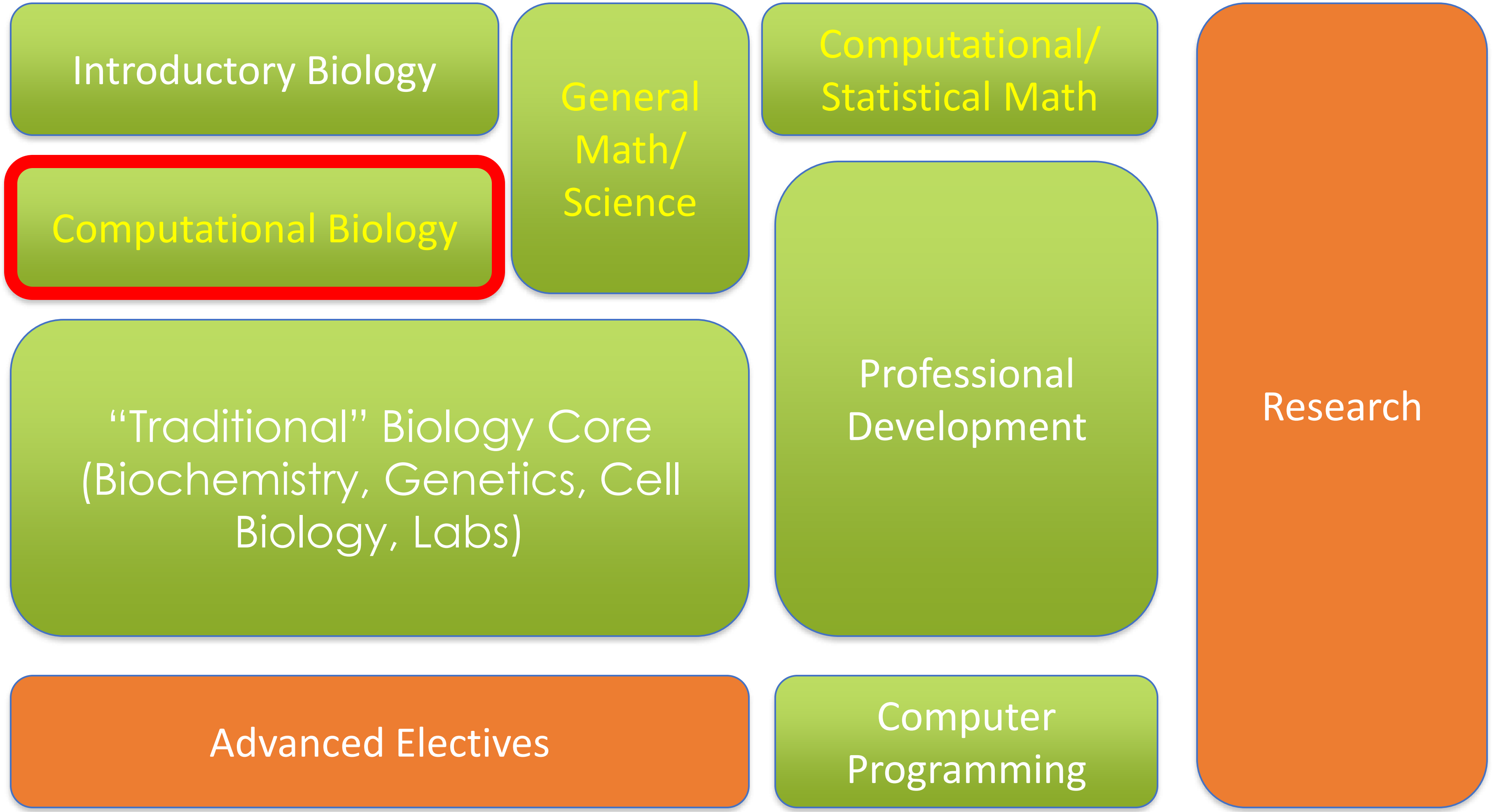
Fitting Computational Biology in an Undergraduate Biology Curriculum: Ten Years Ago





Fitting Computational Biology in an Undergraduate Biology Curriculum: Ten Years Ago





Did We Learn Anything We Could Share? Can We Learn from Others with Similar Needs?

OPEN ACCESS Freely available online PLoS COMPUTATIONAL BIOLOGY

Message from ISCB

A Report of the Curriculum Task Force of the ISCB Education Committee

Lonnie R. Welch[1]*, Russell Schwartz[2]*, Fran Lewitter[3]*

OPEN ACCESS Freely available online PLOS COMPUTATIONAL BIOLOGY

Message from ISCB

Bioinformatics Curriculum Guidelines: Toward a Definition of Core Competencies

Lonnie Welch[1]*, Fran Lewitter[2], Russell Schwartz[3], Cath Brooksbank[4], Predrag Radivojac[5], Bruno Gaeta[6], Maria Victoria Schneider[7]

MESSAGE FROM ISCB

Applying, Evaluating and Refining Bioinformatics Core Competencies (An Update from the Curriculum Task Force of ISCB's Education Committee)

Lonnie Welch[1]*, Cath Brooksbank[2], Russell Schwartz[3], Sarah L. Morgan[2], Bruno Gaeta[4], Alastair M. Kilpatrick[5], Daniel Mietchen[6], Benjamin L. Moore[7], Nicola Mulder[8], Mark Pauley[9], William Pearson[10], Predrag Radivojac[11], Naomi Rosenberg[12], Anne Rosenwald[13], Gabriella Rustici[14], Tandy Warnow[15]

MESSAGE FROM ISCB

The development and application of bioinformatics core competencies to improve bioinformatics training and education

Nicola Mulder[1‡]*, Russell Schwartz[2‡], Michelle D. Brazas[3], Cath Brooksbank[4], Bruno Gaeta[5], Sarah L. Morgan[4], Mark A. Pauley[6], Anne Rosenwald[7], Gabriella Rustici[8], Michael Sierk[9], Tandy Warnow[10], Lonnie Welch[11]

H3ABioNet Pan African Bioinformatics Network for H3Africa | elixir | iSCB INTERNATIONAL SOCIETY FOR COMPUTATIONAL BIOLOGY

H3ABioNet, ELIXIR, GOBLET & the ISCB Education Committee are running a

BIOINFORMATICS EDUCATION SUMMIT

2nd Bioinformatics Education Summit 2020

19th – 22nd May 2020 Virtual event
(Planned to be hosted at EMBL-EBI, Hinxton)

Aim: Bring together Bioinformatics trainers and educators to drive the development of standards and guidelines for Bioinformatics training and education globally

EMBL-EBI | iSCB | elixir | CABANA | H3ABioNet

What Do Students of Computational Biology Need to Know?: ISCB Education Committee Curriculum Task Force

OPEN ACCESS Freely available online
PLOS COMPUTATIONAL BIOLOGY

Message from ISCB

A Report of the Curriculum Task Force of the ISCB Education Committee

Lonnie R. Welch[1]*, Russell Schwartz[2]*, Fran Lewitter[3]*

OPEN ACCESS Freely available online
PLOS COMPUTATIONAL BIOLOGY

Message from ISCB

Bioinformatics Curriculum Guidelines: Toward a Definition of Core Competencies

Lonnie Welch[1]*, Fran Lewitter[2], Russell Schwartz[3], Cath Brooksbank[4], Predrag Radivojac[5], Bruno Gaeta[6], Maria Victoria Schneider[7]

Surveys of employment ads and core facilities directors:

Skill Category	Specific Skills
General	time management, project management, management of multiple projects, independence, curiosity, self-motivation, ability to synthesize information, ability to complete projects, leadership, critical thinking, dedication, ability to communicate scientific concepts, analytical reasoning, scientific creativity, collaborative ability
Computational	programming, software engineering, system administration, algorithm design and analysis, machine learning, data mining, database design and management, scripting languages, ability to use scientific and statistical analysis software packages, open source software repositories, distributed and high-performance computing, networking, web authoring tools, web-based user interface implementation technologies, version control tools
Biology	molecular biology, genomics, genetics, cell biology, biochemistry, evolutionary theory, regulatory genomics, systems biology, next generation sequencing, proteomics/mass spectrometry, specialized knowledge in one or more domains
Statistics and Mathematics	application of statistics in the contexts of molecular biology and genomics, mastery of relevant statistical and mathematical modeling methods (including experimental design, descriptive and inferential statistics, probability theory, differential equations and parameter estimation, graph theory, epidemiological data analysis, analysis of next generation sequencing data using R and Bioconductor)
Bioinformatics	analysis of biological data; working in a production environment managing scientific data; modeling and warehousing of biological data; using and building ontologies; retrieving and manipulating data from public repositories; ability to manage, interpret, and analyze large data sets; broad knowledge of bioinformatics analysis methodologies; familiarity with functional genetic and genomic data; expertise in common bioinformatics software packages, tools, and algorithms

doi:10.1371/journal.pcbi.1003496.t001

What Do Computational Biologists Need to Know?: the ISCB Education Committee Curriculum Task Force

OPEN ACCESS Freely available online

PLOS COMPUTATIONAL BIOLOGY

Message from ISCB

A Report of the Curriculum Task Force of the ISCB Education Committee

Lonnie R. Welch[1]*, Russell Schwartz[2]*, Fran Lewitter[3]*

Surveys of degree programs (PhD, MS, BS, and Certificate) in Computational Biology and Bioinformatics:

a. Computational Biology
- a.1. Computational molecular biology/Genetics
- a.2. Computational structural biology
- a.3. Biological simulation
- a.4. Bioimage analysis
- a.5. Computational Biology – General
- a.6. Computational Biology – Other (please specify)

b. Computer Science
- b.1. Programming
- b.2. Software engineering
- b.3. Algorithms/Data Structures
- b.4. Databases
- b.5. Artificial Intelligence
- b.6. Machine Learning
- b.7. Visualization
- b.8. Computer Systems
- b.9. Computer Science – Other (please specify)

c. Biology
- c.1. Genetics
- c.2. Cell biology
- c.3. Biochemistry
- c.4. Biophysics
- c.5. Evolutionary biology
- c.6. Biotechnology
- c.7. Genomics
- c.8. Systems biology
- c.9. Molecular biology
- c.10. Biology laboratory
- c.11. Biology – General
- c.12. Biology – Other (please specify)

d. Mathematics/Statistics
- d.1. Probability
- d.2. Statistics
- d.3. Biostatistics
- d.4. Mathematical biology
- d.5. Differential equations
- d.6. Linear algebra
- d.7. Discrete mathematics
- d.8. Calculus
- d.9. Mathematics/Statistics – Other (please specify)

e. Other Science
- e.1. Physics
- e.2. Chemistry
- e.3. Science – Other (please specify)

f. Non-Science Coursework
- f.1. Ethics
- f.2. Entrepreneurship
- f.3. Writing/Communications
- f.4. Non-science – Other (please specify)

g. Other Requirements
- g.1. Seminar
- g.2. Research
- g.3. Internship
- g.4. Capstone project
- g.5. Teaching assistantship
- g.6. Other requirement not listed (please specify)

h. Other Elective
- h.1. Elective (possibly restricted)

What Do Computational Biologists Need to Know?: the ISCB Curriculum Task Force

OPEN ACCESS Freely available online PLOS COMPUTATIONAL BIOLOGY

Message from ISCB

A Report of the Curriculum Task Force of the ISCB Education Committee

Lonnie R. Welch[1]*, Russell Schwartz[2]*, Fran Lewitter[3]*

OPEN ACCESS Freely available online PLOS COMPUTATIONAL BIOLOGY

Message from ISCB

Bioinformatics Curriculum Guidelines: Toward a Definition of Core Competencies

Lonnie Welch[1]*, Fran Lewitter[2], Russell Schwartz[3], Cath Brooksbank[4], Predrag Radivojac[5], Bruno Gaeta[6], Maria Victoria Schneider[7]

Conclusion: There is little agreement among degree programs and little consensus between employers and instructors. Therefore:

- We could provide some framework for understanding the space of programs and needs;
- but, it was premature to be prescriptive about best practices;
- and, we needed to learn more and get input from the community.

A First Pass at Defining Core Competencies (ISCB Core Competencies v1.0)

OPEN ACCESS Freely available online

PLOS COMPUTATIONAL BIOLOGY

Message from ISCB

Bioinformatics Curriculum Guidelines: Toward a Definition of Core Competencies

Lonnie Welch[1]*, Fran Lewitter[2], Russell Schwartz[3], Cath Brooksbank[4], Predrag Radivojac[5], Bruno Gaeta[6], Maria Victoria Schneider[7]

(a) An ability to apply knowledge of computing, biology, statistics, and mathematics appropriate to the discipline.

(b) An ability to analyze a problem and identify and define the computing requirements appropriate to its solution.

(c) An ability to design, implement, and evaluate a computer-based system, process, component, or program to meet desired needs in scientific environments.

(d) An ability to use current techniques, skills, and tools necessary for computational biology practice.

(e) An ability to apply mathematical foundations, algorithmic principles, and computer science theory in the modeling and design of computer-based systems in a way that demonstrates comprehension of the tradeoffs involved in design choices.

(f) An ability to apply design and development principles in the construction of software systems of varying complexity.

(g) An ability to function effectively on teams to accomplish a common goal.

(h) An understanding of professional, ethical, legal, security, and social issues and responsibilities.

(i) An ability to communicate effectively with a range of audiences.

(j) An ability to analyze the local and global impact of bioinformatics and genomics on individuals, organizations, and society.

(k) Recognition of the need for and an ability to engage in continuing professional development.

(l) Detailed understanding of the scientific discovery process and of the role of bioinformatics in it.

(m) An ability to apply statistical research methods in the contexts of molecular biology, genomics, medical, and population genetics research.

(n) Knowledge of general biology, in-depth knowledge of at least one area of biology, and understanding of biological data generation technologies.

Mapping Competencies to Personas

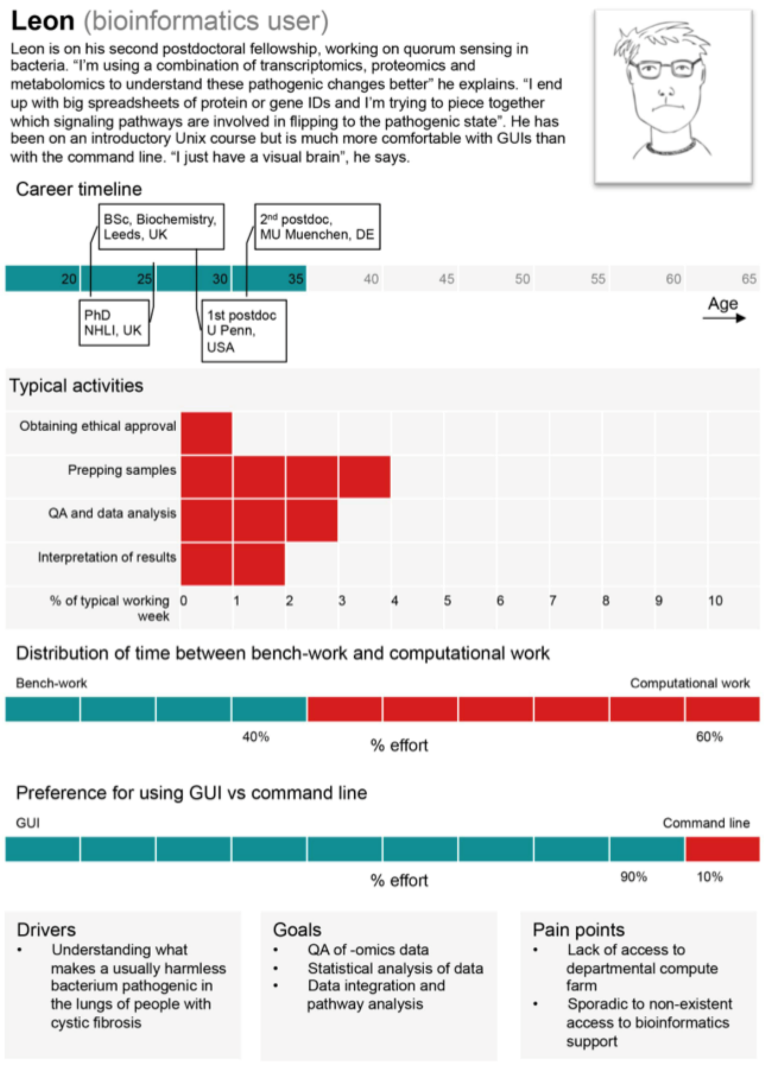


Figure 2. A persona based on a typical "bioinformatics user." QA: Quality Assurance, GUI: Graphical User Interface. Image credit: Jenny Cham, Mary Todd Bergman, and Cath Brooksbank, EMBL-EBI.
doi:10.1371/journal.pcbi.1003496.g002

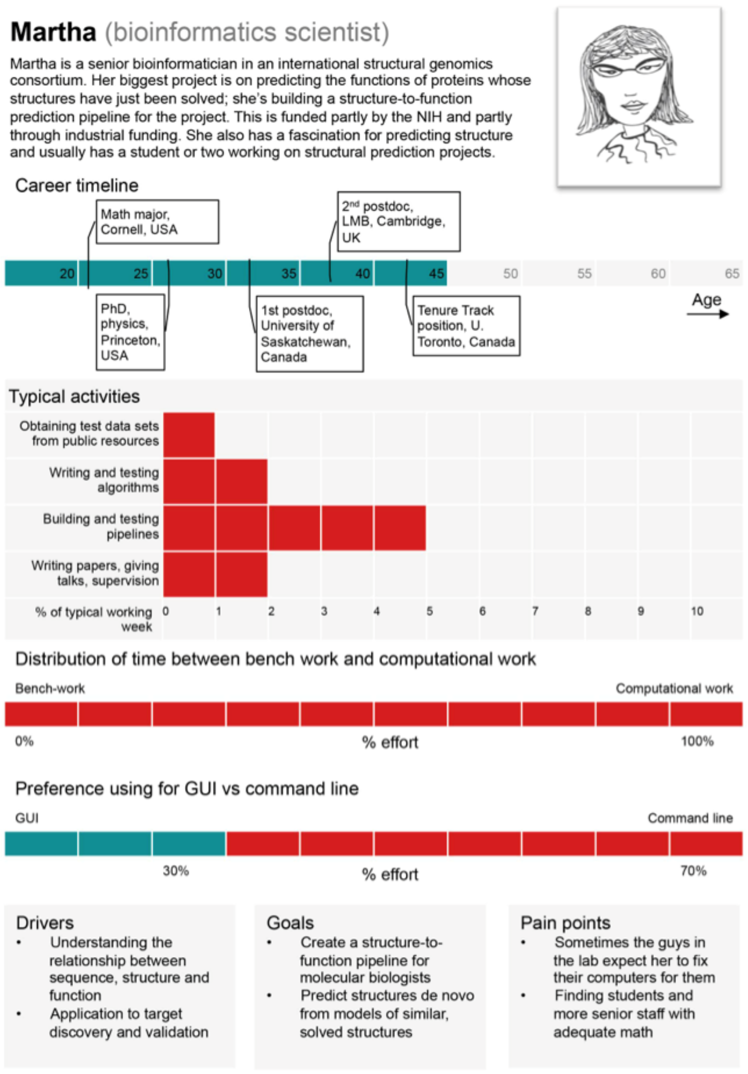


Figure 3. A persona based on a typical "bioinformatics scientist." GUI: Graphical User Interface. Image credit: Jenny Cham, Mary Todd Bergman, and Cath Brooksbank, EMBL-EBI.
doi:10.1371/journal.pcbi.1003496.g003

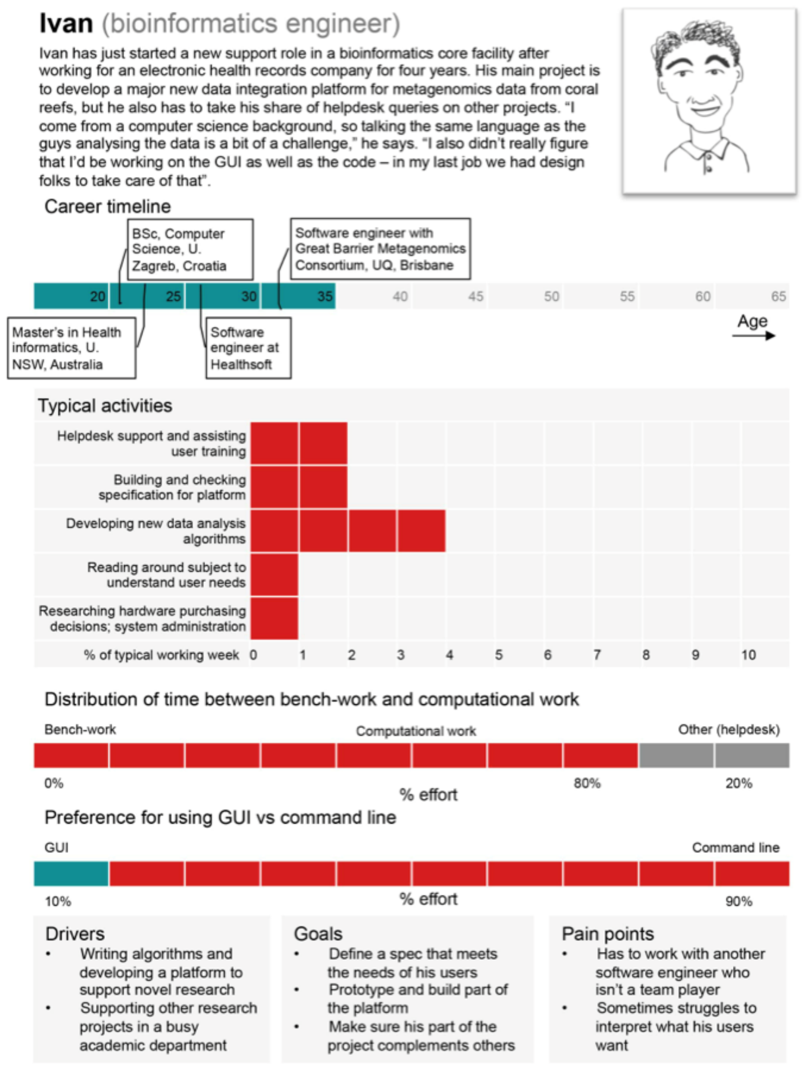
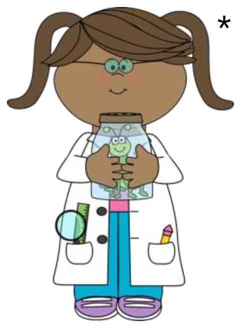


Figure 4. A persona based on a typical "bioinformatics engineer." GUI: Graphical User Interface. Image credit: Jenny Cham, Mary Todd Bergman, and Cath Brooksbank, EMBL-EBI.
doi:10.1371/journal.pcbi.1003496.g004

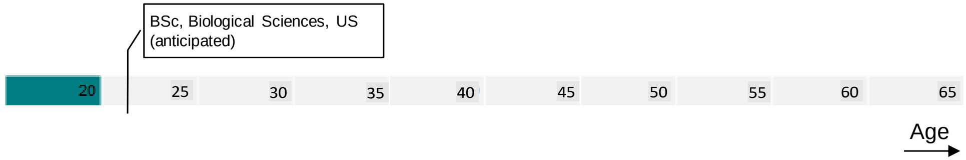
Example Persona: A Bioinformatics User

Angela (bioinformatics user)

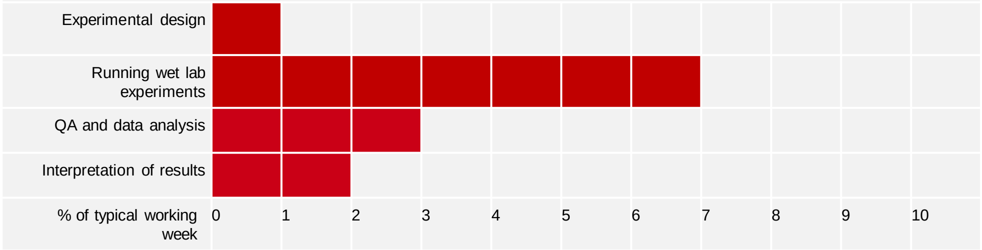


Angela is an undergraduate biology student who is still weighing a few possible career paths. "I always planned to get my M.D. and work in medicine, but I really love research and am considering a Ph.D. in Biology and a career in research. I might take a post- baccalaureate year to get more experience and weigh my options." Her primary interest is experimental biology, but she has come to appreciate how important data analysis methods are to her work. "I enjoy working with my hands and planning and running experiments, but I appreciate how much this work depends on statistical analysis and data science. I want the skills that will get me into a good job or degree program right now, but I want to be prepared for what biology and medicine will look like 30 or 40 years from now."

Career timeline



Typical activities



Distribution of time between bench-work and computational work

Preference for using GUI vs command line

Drivers

- Gain experience in coursework and research that will prepare her for a variety of career options in biology or medicine

Goals

- Learn how to design and run experiments
- Apply statistical/data analysis methods
- Data integration and pathway analysis

Pain points

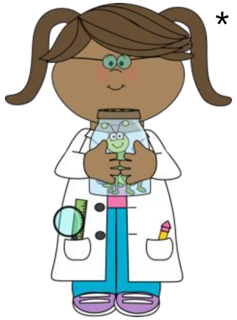
- Limited prior background in computing and mathematics
- Limited access to mentorship on these topics in her labs

*Image "Girl Scientist with Insect Jar Clip Art" from MyCuteGraphics.com

Example Persona: A Bioinformatics User

Angela (bioinformatics user)

Angela is an undergraduate biology student who is still weighing a few possible career paths. “I always planned to get my M.D. and work in medicine, but I really love research and am considering a Ph.D. in Biology and a career in research. I might take a post- baccalaureate year to get more experience and weigh my options.” Her primary interest is experimental biology, but she has come to appreciate how important data analysis methods are to her work. “I enjoy working with my hands and planning and running experiments, but I appreciate how much this work depends on statistical analysis and data science. I want the skills that will get me into a good job or degree program right now, but I want to be prepared for what biology and medicine will look like 30 or 40 years from now.”


*

Career timeline

Angela is an undergraduate biology student who is still weighing a few possible career paths. “I always planned to get my M.D. and work in medicine, but I really love research and am considering a Ph.D. in Biology and a career in research. I might take a post- baccalaureate year to get more experience and weigh my options.” Her primary interest is experimental biology, but she has come to appreciate how important data analysis methods are to her work. “I enjoy working with my hands and planning and running experiments, but I appreciate how much this work depends on statistical analysis and data science. I want the skills that will get me into a good job or degree program right now, but I want to be prepared for what biology and medicine will look like 30 or 40 years from now.”

Preference for using GUI vs command line

Drivers

- Gain experience in coursework and research that will prepare her for a variety of career options in biology or medicine

Goals

- Learn how to design and run experiments
- Apply statistical/data analysis methods
- Data integration and pathway analysis

Pain points

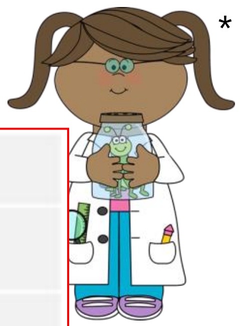
- Limited prior background in computing and mathematics
- Limited access to mentorship on these topics in her labs

*Image “Girl Scientist with Insect Jar Clip Art” from MyCuteGraphics.com

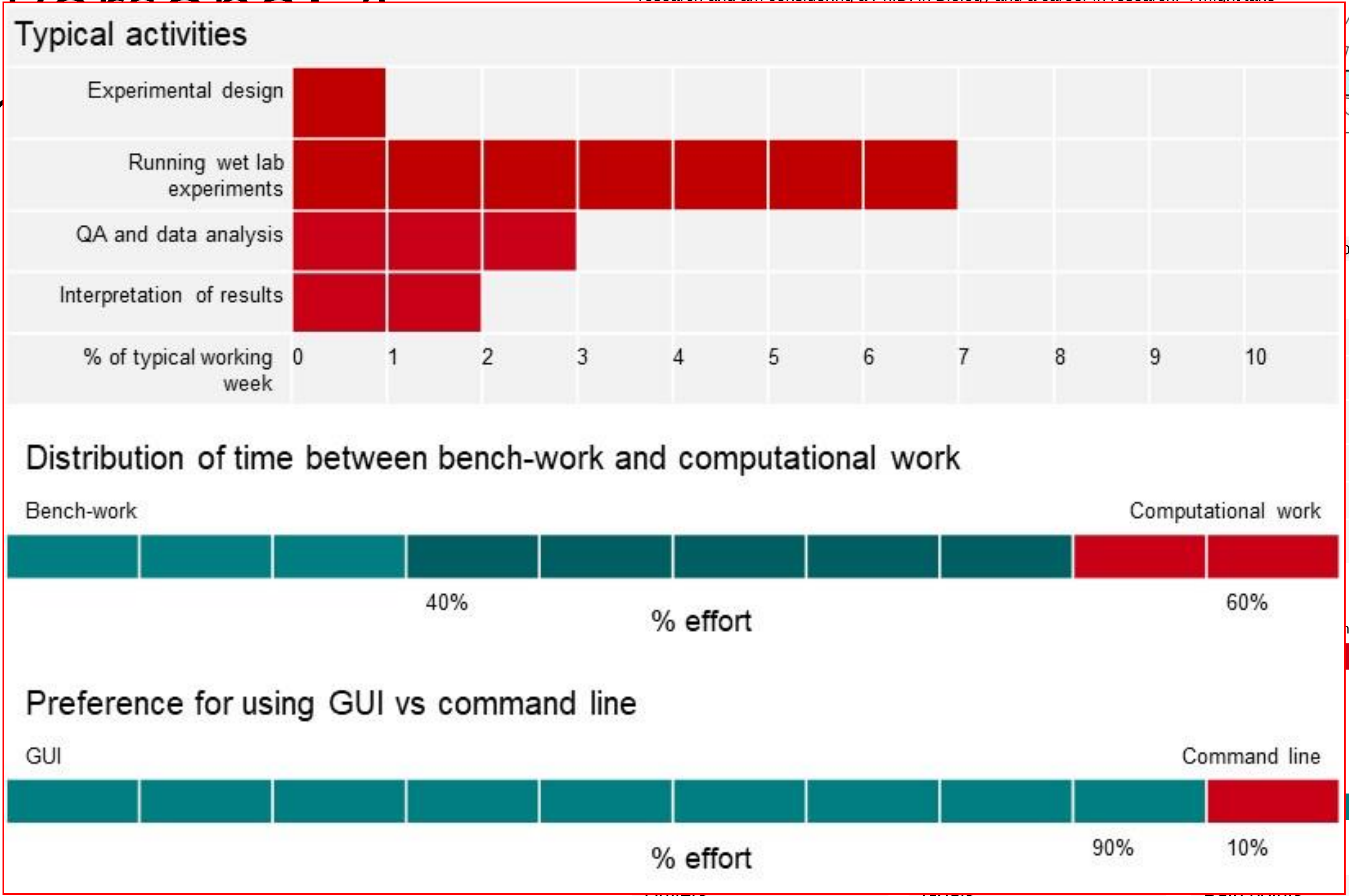
Example Persona: A Bioinform

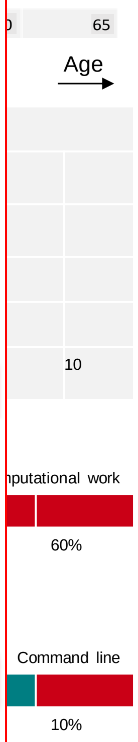
Angela (bioinformatics user)

Angela is an undergraduate biology student who is still weighing a few possible career paths. "I always planned to get my M.D. and work in medicine, but I really love research and am considering a Ph.D. in Biology and a career in research. I might take



*



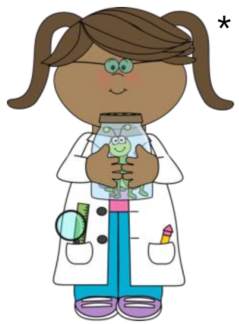


Drivers	Goals	Pain points
• Gain experience in coursework and research that will prepare her for a variety of career options in biology or medicine	• Learn how to design and run experiments • Apply statistical/data analysis methods • Data integration and pathway analysis	• Limited prior background in computing and mathematics • Limited access to mentorship on these topics in her labs

*Image "Girl Scientist with Insect Jar Clip Art" from MyCuteGraphics.com

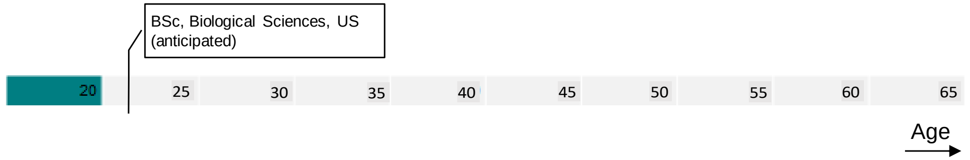
Example Persona: A Bioinformatics User

Angela (bioinformatics user)

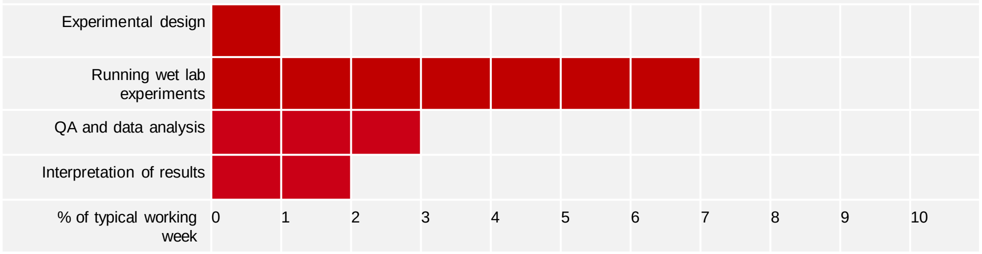


Angela is an undergraduate biology student who is still weighing a few possible career paths. "I always planned to get my M.D. and work in medicine, but I really love research and am considering a Ph.D. in Biology and a career in research. I might take a post- baccalaureate year to get more experience and weigh my options." Her primary interest is experimental biology, but she has come to appreciate how important data analysis methods are to her work. "I enjoy working with my hands and planning and running experiments, but I appreciate how much this work depends on statistical analysis and data science. I want the skills that will get me into a good job or degree program right now, but I want to be prepared for what biology and medicine will look like 30 or 40 years from now."

Career timeline



Typical activities



Distribution of time between bench-work and computational work

Preference for using GUI vs command line

Drivers
- Gain experience in coursework and research that will prepare her for a variety of career options in biology or medicine

Goals
- Learn how to design and run experiments
- Apply statistical/data analysis methods
- Data integration and pathway analysis

Pain points
- Limited prior background in computing and mathematics
- Limited access to mentorship on these topics in her labs

*Image "Girl Scientist with Insect Jar Clip Art" from MyCuteGraphics.com

Phase II: Getting to a Community Consensus

The development and application of bioinformatics core competencies to improve bioinformatics training and education

Nicola Mulder[1‡*], Russell Schwartz[2‡], Michelle D. Brazas[3], Cath Brooksbank[4], Bruno Gaeta[5], Sarah L. Morgan[4], Mark A. Pauley[6], Anne Rosenwald[7], Gabriella Rustici[8], Michael Sierk[9], Tandy Warnow[10], Lonnie Welch[11]

- We conducted many workshops at ISMB and ISCB- and Goblet-affiliated meetings to solicit feedback and refine our notion of competencies
- Results:
    - Refined competency framework incorporating Bloom's Taxonomy to capture varying levels of skill
    - Expanded to a broader set of personas to better capture variation across training needs
    - Gathered experiences on how competencies were used to develop and evaluate programs

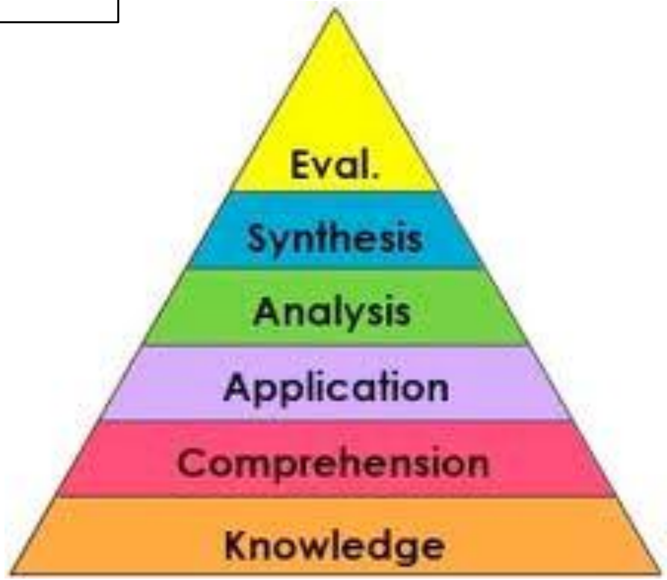


Image from Wikimedia Commons
File:Bloom taxonomy.jpg, user Xristina la

ISCB Competencies v2.0

Competency \ Persona
A. General biology
B. Depth in at least one area of biology (e.g., evolutionary biology, genetics, molecular biology, biochemistry, anatomy, physiology)
C. Biological data generation technologies.
D. Details of the scientific discovery process and of the role of bioinformatics in it.
E. Statistical research methods in the context of molecular biology, genomics, medical, and population genetics research.
F. Bioinformatics tools and their usage.
G. The ability of a computer-based system, process, algorithm, component, or program to meet desired needs in scientific environments/problem.
H. Computing requirements appropriate to solve a given scientific problem (e.g. system, process, algorithm, component or program; define algorithmic time and space complexities and hardware resources required to solve a problem).

I. GUI/Web-based computing skills appropriate to the discipline (e.g., effectively use bioinformatics and analysis tools through web).
J. Command line and scripting-based computing skills appropriate to the discipline.
K. Construction of software systems of varying complexity based on design and development principles.
L. Local and global impact of bioinformatics and genomics on individuals, organizations, and society.
M. Professional, ethical, legal, security and social issues and responsibilities of bioinformatics and genomic data in the workplace.
N. Effective communication of bioinformatics and genomics problem/issue/topics with a range of audiences, including, but not limited to, other bioinformatics professionals
O. Effective teamwork to accomplish a common scientific goal.
P. Engage in continuing professional development in bioinformatics.
https://doi.org/10.1371/journal.pcbi.1005772.t003

ISCB Competencies v2.0

I. GUI/Web-based computing skills appropriate to the discipline

Competency \ Persona	Discovery biologist/ academic life science researcher	Molecular life science educator	Academic bioinformatics researcher
A. General biology	evaluation	comprehension	synthesis
B. Depth in at least one area of biology (e.g., evolutionary biology, genetics, molecular biology, biochemistry, anatomy, physiology)	evaluation	analysis	evaluation
C. Biological data generation technologies.	evaluation	understand	evaluation
D. Details of the scientific discovery process and of the role of bioinformatics in it.	application	evaluation	synthesis to evaluation
E. Statistical research methods in the context of molecular biology, genomics, medical, and population genetics research.	application	evaluation	synthesis to evaluation
F. Bioinformatics tools and their usage.	application	evaluation	synthesis to evaluation
G. The ability of a computer-based system, process, algorithm, component, or program to meet desired needs in scientific environments/problem.	application	comprehension	synthesis to evaluation
H. Computing requirements appropriate to solve a given scientific problem (e.g. system, process, algorithm, component or program; resources required to solve a problem).	application	comprehension	synthesis to evaluation

Further Refinement: ISCB Competencies v3.0

2nd Bioinformatics Education Summit 2020

19th – 22nd May 2020 Virtual event
(Planned to be hosted at EMBL-EBI, Hinxton)

Aim: Bring together Bioinformatics trainers and educators to drive the development of standards and guidelines for Bioinformatics training and education globally

EMBL-EBI iSCB elixir CABANA H3ABioNet

- Revised collection of competencies
- Revised assignment to an expanded list of personas
- Refinement in terms of Knowledge, Skills, and Attitudes (KSAs)
    - Knowledge: what do you need to know to exhibit this competency
    - Skills: what do you need to be able to do to exhibit this competency
    - Attributes: how does a person with this competency behave

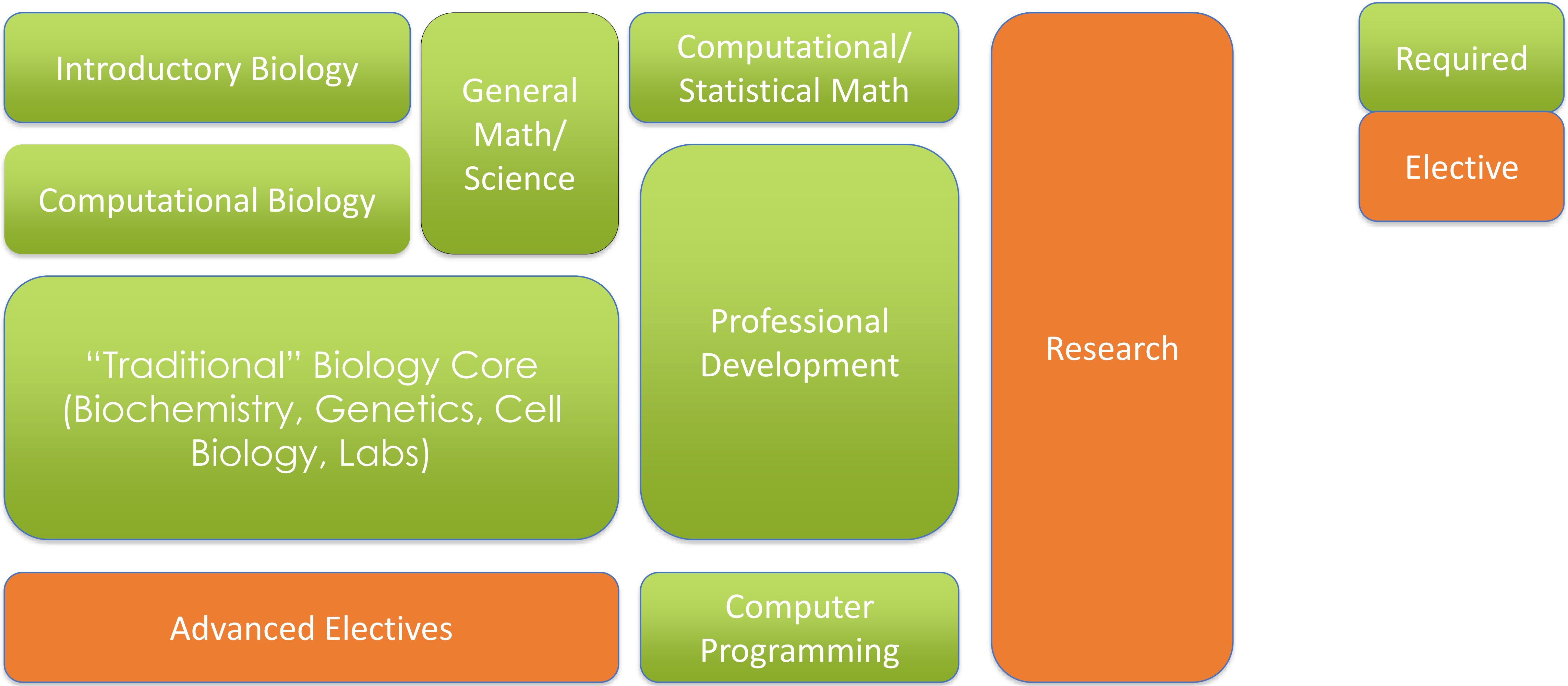
ISCB Competencies v3.0

A. Work at depth in at least one technical area aligned with the life sciences.
B. Prepare life science data for computational analysis
C. Have a positive impact on scientific discovery through bioinformatics
D. Use data science methods suitable for the size and complexity of the data
E. Manage own and others' data according to community standards and principles
F. Make appropriate use of bioinformatics tools and resources
G. Contribute effectively to the design and development of user-centric bioinformatics tools and resources
H. Make appropriate and efficient use of scripting and programming languages
I. Construct, manage and maintain bioinformatics computing infrastructure of varying complexity
J. Comply with professional, ethical, legal and social standards and codes of conduct relevant to computational biology
K. Communicate meaningfully with a range of audiences - within and beyond your profession
L. Work effectively in teams to accomplish a common goal
M. Engage in continuing professional development in bioinformatics

A. Work at depth in at least one technical area aligned with the life sciences.

What do you need to know to exhibit competency in this area?	**What skills do you need to exhibit competency in this area?**
KA3-1. The central dogma, general biological concepts and how they relate to each other. KA3-2 (UA K1). A technical area aligned with the life sciences, and the core experimental platform or data-generating technologies in the chosen field. KA3-3. Tools and databases relevant to the specific topic area. KA3-4. Details of the experimental process.	SA3-1. Differentiates between biological and non-biological entities. SA3-2. Relays biological information and critiques life science papers. SA3-3. Plans and performs experiments responsibly. SA3-4. Asks relevant, hypothesis driven, well-defined biological questions and accurately judges the validity of results. SA3-5. Interprets experimental results appropriately and sets them in the context of broader knowledge.
How does a person with this competence behave?	**How does a person with this competence avoid behaving?**
AA3-2. Integrates ideas from the broader scientific community. AA3-3. Embraces collaboration in the field. AA3-4. Considers the broader impact of research in the field. AA3-5. Demonstrates critical and creative thinking and encourages it in others.	NA3-1. Is narrow minded about the boundaries of the field of biology. NA3-2. Defends the discipline to the exclusion of others.

Competency framework provides a basis to revisit program design & assessment



Introduction to Computational Biology: Official Learning Objectives

1. Learn major biological data types, the methods by which they are produced, and their uses.
2. Learn to critically assess the reliability of biological data sources.
3. Learn essential concepts of statistics and algorithms needed to productively use database search, analysis, and inference tools and interpret their results.
4. Learn to synthesize results from different data sources and select sources appropriate to a given problem.
5. Learn about of the major repositories of biological data and the tools to access them.
6. Learn to independently research a biological question using online resources.
7. Learn how to pose biological questions through mathematical models and reason about the assumptions and limitations of those models.
8. Learn to simulate the behavior of simple mathematical models.
9. Learn basic tools and concepts of biological image analysis.

V2 Competencies

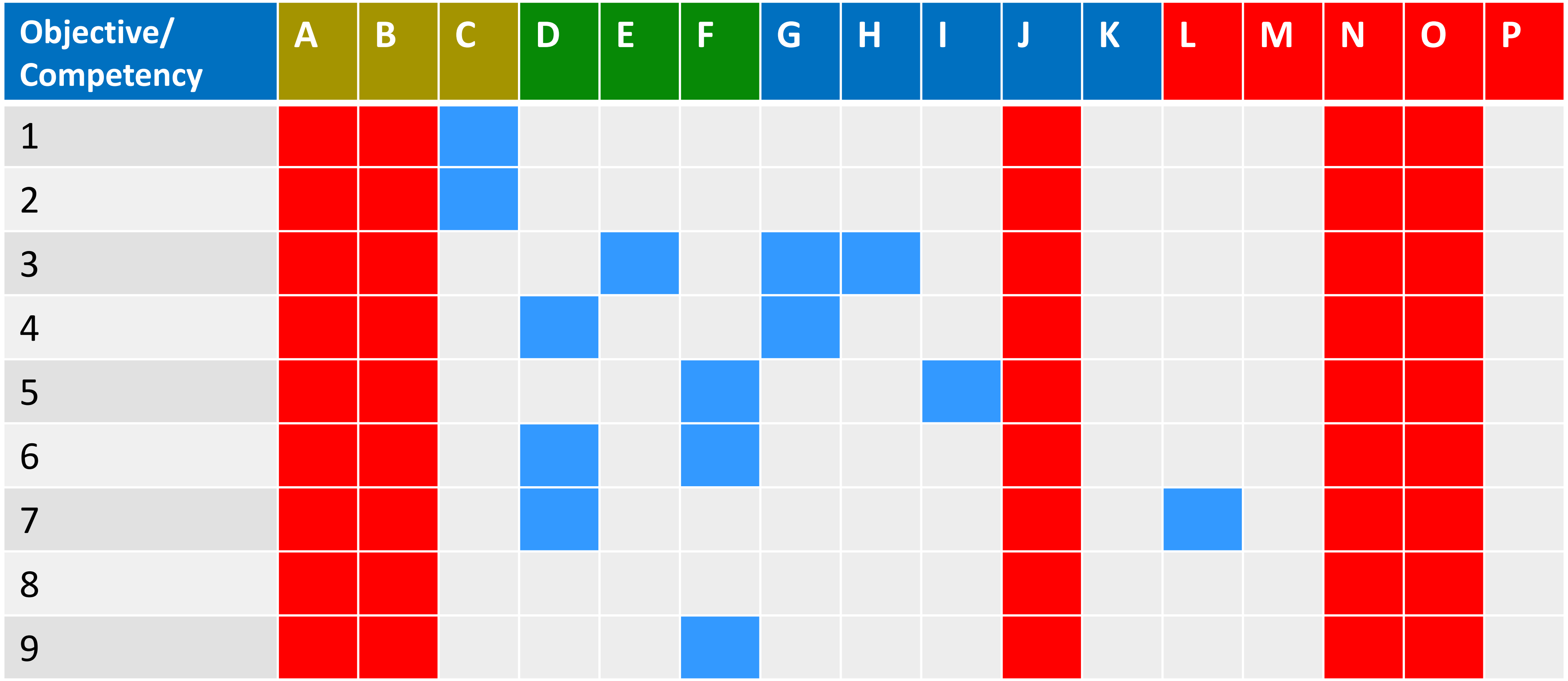
	Competency
BIOLOGY	**A.** General biology
	B. Depth in at least one area of biology (e.g., evolutionary biology, genetics, molecular biology, biochemistry, anatomy, physiology)
	C. Biological data generation technologies.
COMPUTATIONAL BIOLOGY	**D.** Details of the scientific discovery process and of the role of bioinformatics in it.
	E. Statistical research methods in the context of molecular biology, genomics, medical, and population genetics research.
	F. Bioinformatics tools and their usage.
COMPUTER SCIENCE	**G.** The ability of a computer-based system, process, algorithm, component, or program to meet desired needs in scientific environments/problem.
	H. Computing requirements appropriate to solve a given scientific problem (e.g. system, process, algorithm, component or program; define algorithmic time and space complexities and hardware resources required to solve a problem).
	I. GUI/Web based computing skills appropriate to the discipline (e.g., effectively use bioinformatics and analysis tools through web).
	J. Command line and scripting based computing skills appropriate to the discipline.
	K. Construction of software systems of varying complexity based on design and development principles.
PROFESSIONAL COMPETENCIES	**L.** Local and global impact of bioinformatics and genomics on individuals, organizations, and society.
	M. Professional, ethical, legal, security and social issues and responsibilities of bioinformatics and genomic data in the workplace.
	N. Effective communication of bioinformatics and genomics problem/issue/topics with a range of audiences, including, but not limited to, other bioinformatics professionals
	O. Effective team work to accomplish a common scientific goal.
	P. Engage in continuing professional development in bioinformatics.

Comparing to ISCB EduComm Core Competencies

Objective/ Competency	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1																
2																
3																
4																
5																
6																
7																
8																
9																

Material covered in Intro to Computational Biology

Comparing to ISCB EduComm Core Competencies



Material covered in Intro to Computational Biology

Material covered elsewhere in our curriculum.

Comparing to ISCB EduComm Core Competencies

Objective/ Competency	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1																
2																
3																
4																
5																
6																
7																
8																
9																

Material covered in Intro to Computational Biology

Material covered elsewhere in our curriculum.

Material lacking from our undergrad biology training.

BIOLOGY	
COMPUTATIONAL BIOLOGY	
	Competency
COMPUTER SCIENCE	**K.** Construction of software systems of varying complexity based on design and development principles.
PROFESSIONAL COMPETENCIES	**M.** Professional, ethical, legal, security and social issues and responsibilities of bioinformatics and genomic data in the workplace.
	P. Engage in continuing professional development in bioinformatics.

Fitting Computational Biology in an Undergraduate Biology Curriculum: The Future?

Required

Elective

Introductory Biology

Computer Programming

Computational/ Statistical Math

Computational Biology

General Math/Science

Professional Development

Research

"Traditional" Biology Core (Biochemistry, Genetics, Cell Biology, Labs)

Advanced Electives

Training Ten Years Later: Computational Education of Biology PhD Students

Coursework

Research Rotations

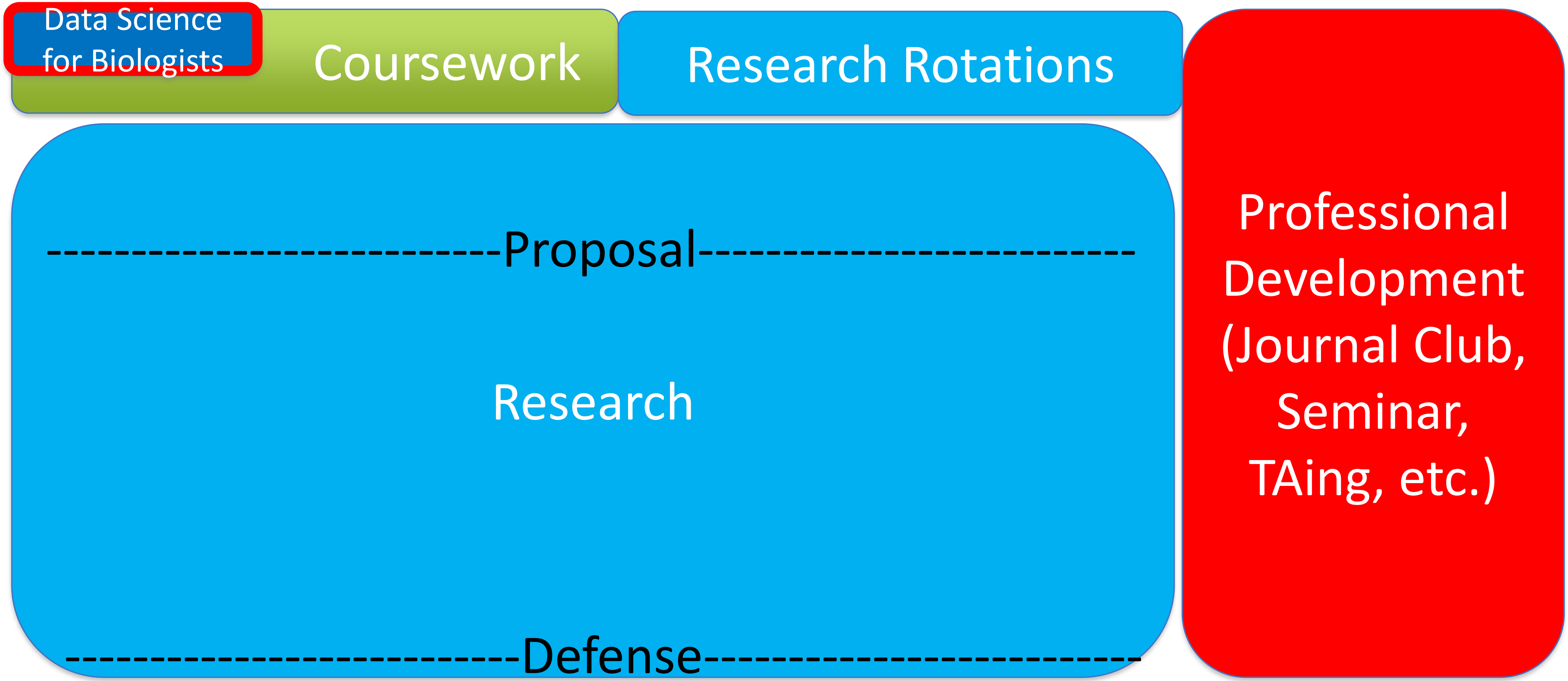
---------------------------Proposal---------------------------

Research

---------------------------Defense---------------------------

Professional Development (Journal Club, Seminar, TAing, etc.)

Designing Computational Training for Biology PhD Students



Thinking About the Persona of a Biology PhD Student

ISCB Competency V3	**Discovery biologist**	**Molecular life science educator**
General biology	Evaluation	Comprehension
Depth in at least one area of biology (e.g., evolutionary biology, genetics, molecular biology, biochemistry, anatomy, physiology)	Evaluation	Analysis
Biological data generation technologies.	Evaluation	Understanding
Details of the scientific discovery process and of the role of bioinformatics in it.	Application	Evaluation
Statistical, machine learning and data science research methods in the context of molecular biology, genomics, medical, and population genetics research.	Application	Evaluation
Data management	N/A	N/A
Bioinformatics tools and resources and their usage.	Application	Evaluation
Fundamentals of computer science systems	Application	Comprehension
Fundamentals of computer science theory	Application	Comprehension
Human–computer interaction (HCI)	Application	Comprehension
Scripting and programming appropriate to the discipline	Application	Comprehension
Construction of bioinformatics computing infrastructure (software and/or hardware) of varying complexity based on design and development principles	Comprehension	Comprehension
Local and global impact of bioinformatics and genomics on individuals, organizations, and society.	Knowledge	Comprehension
Professional, ethical, legal, security and social issues and responsibilities of bioinformatics and genomic data in the workplace, including non-research settings	Application	Comprehension
Effective communication of bioinformatics and genomics problem/issue/topics with a range of audiences, including, but not limited to, other bioinformatics professionals	Application	Comprehension
Effective teamwork to accomplish a common scientific goal.	Application	Analysis
Engage in continuing professional development in bioinformatics.	Application	Application

ISCB Competencies v3.0

A. Work at depth in at least one technical area aligned with the life sciences.
B. Prepare life science data for computational analysis
C. Have a positive impact on scientific discovery through bioinformatics
D. Use data science methods suitable for the size and complexity of the data
E. Manage own and others' data according to community standards and principles
F. Make appropriate use of bioinformatics tools and resources
G. Contribute effectively to the design and development of user-centric bioinformatics tools and resources
H. Make appropriate and efficient use of scripting and programming languages
I. Construct, manage and maintain bioinformatics computing infrastructure of varying complexity
J. Comply with professional, ethical, legal and social standards and codes of conduct relevant to computational biology
K. Communicate meaningfully with a range of audiences - within and beyond your profession
L. Work effectively in teams to accomplish a common goal
M. Engage in continuing professional development in bioinformatics

Goals and Constraints for PhD Students

Students coming out of our class should be able to:

- analyze data from a variety of experimental modalities
- plan experiments with rigorous data analysis in mind
- perform these tasks for a variety of biological domains, tasks, and data sources
- generalize to new hypotheses, data sources, and biological domains

The PhD students cannot:

- have prerequisites, given lack of standard training at the undergrad level
- take many courses or multiple semesters to reach competency for their lab work

Data Science for Biologists: Learning Objectives

1. To identify and apply basic principles of statistical reasoning to biological data analysis.
2. To identify and apply cross-cutting computational and statistical analyses methods commonly used in biological data analysis.
3. To use standard computational data analysis software on biological data sources.
4. To reason about, formulate, and modify study designs to make statistically robust inferences from resulting data.
5. To apply standard data analysis methods and software to common forms of high-throughput biological data.
6. To generalize elements of data science and statistical reasoning to novel problem domains and data types.

Course Structure

Module 1: RNA-seq in Developmental Biology

(w/ Veronica Hinman)

Module 2: DNA-seq in Microbiology

(w/ Luisa Hiller)

Module 3: Neural Imaging

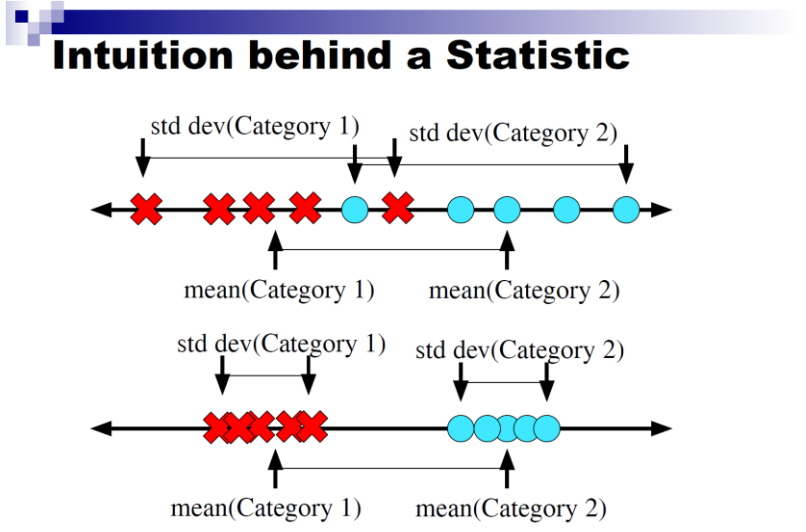
(w/ Alison Barth)

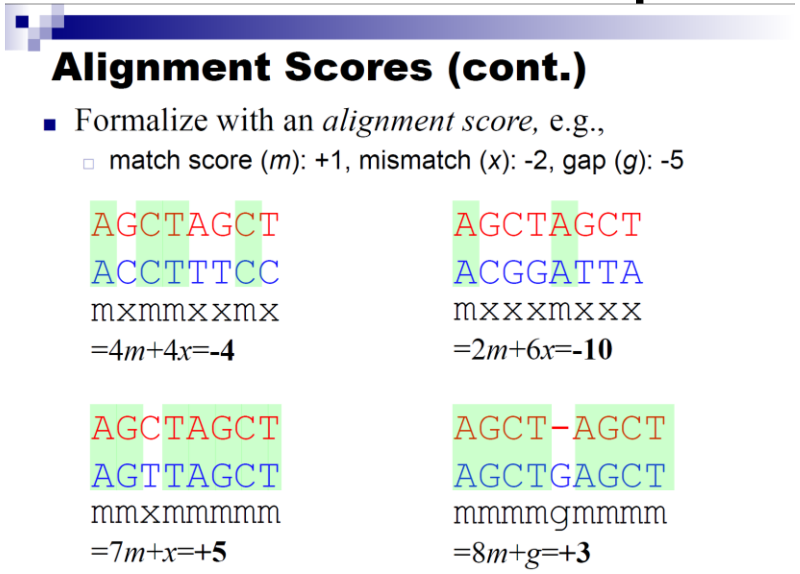
Module 4: Electrophysiology

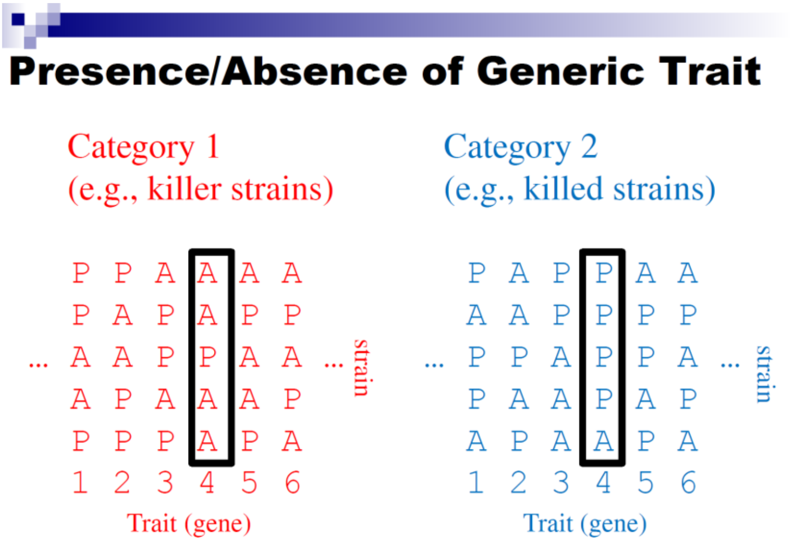
(w/ Sandy Kuhlman)

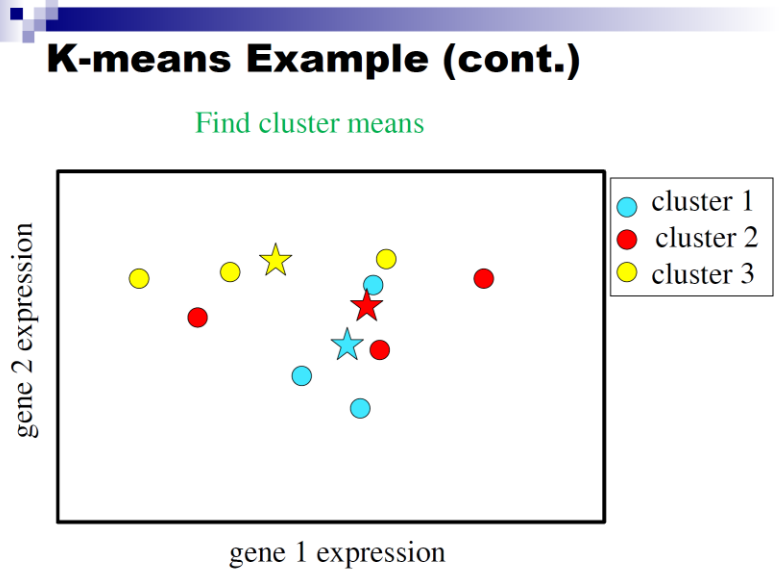
	Day 1	Day 2	Day 3
Module 1 Lectures	Developmental biology and expression analysis	Differential expression analysis and clustering	
Module 1 Workshops	Differential expression analysis for RNA-seq data	Clustering for RNA-seq data	
Module 2 Lectures	DNA sequences and sequence alignment	Microbial communities and bacterocins	Hypothesis testing for categorical data
Module 2 Workshops	Genome annotation by DNA sequence alignment	Identifying orthology groups and differentially conserved genes	
Module 3 Lectures	Neuroscience basics and neural connectivity	Unsupervised classification	
Module 3 Workshops	Hypothesis-driven image feature analysis	Unsupervised classification of image features	
Module 4 Lectures	Electrophysiology basics	Dimensionality reduction	
Module 4 Workshops	Dimensionality reduction from neural recordings	Regression analysis on neural data	

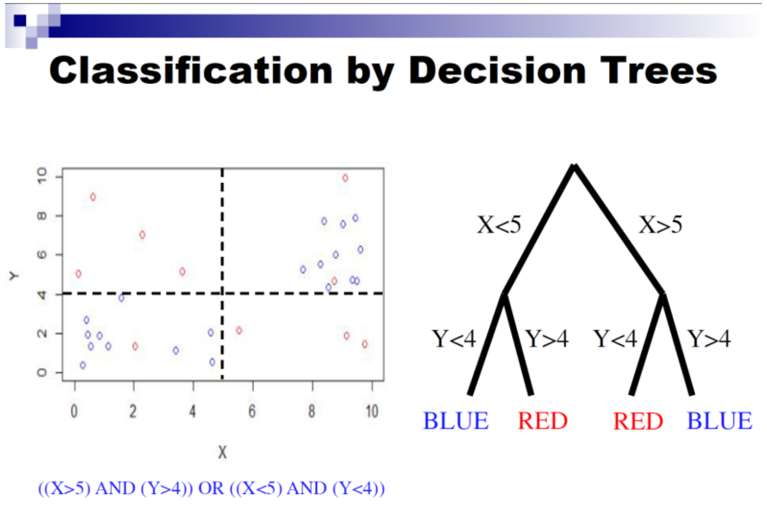
Lectures introduce basic bioinformatics and biostatistics tools and concepts

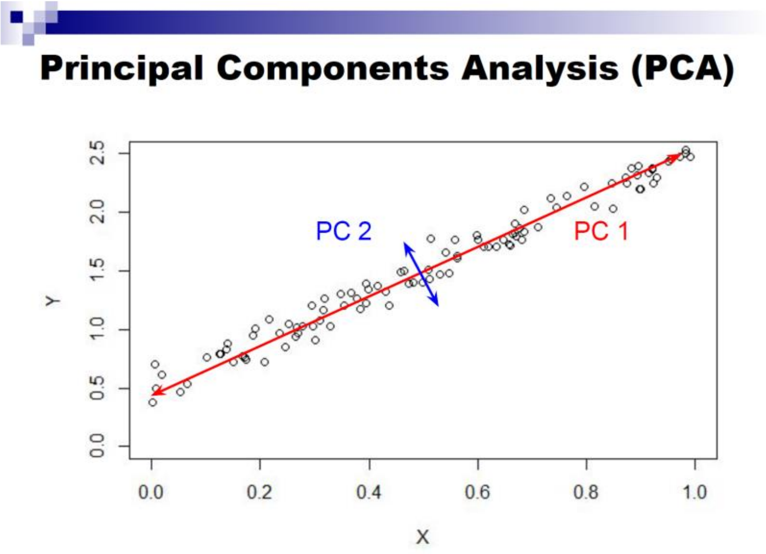












Concepts are reinforced by transfer across domains and data types

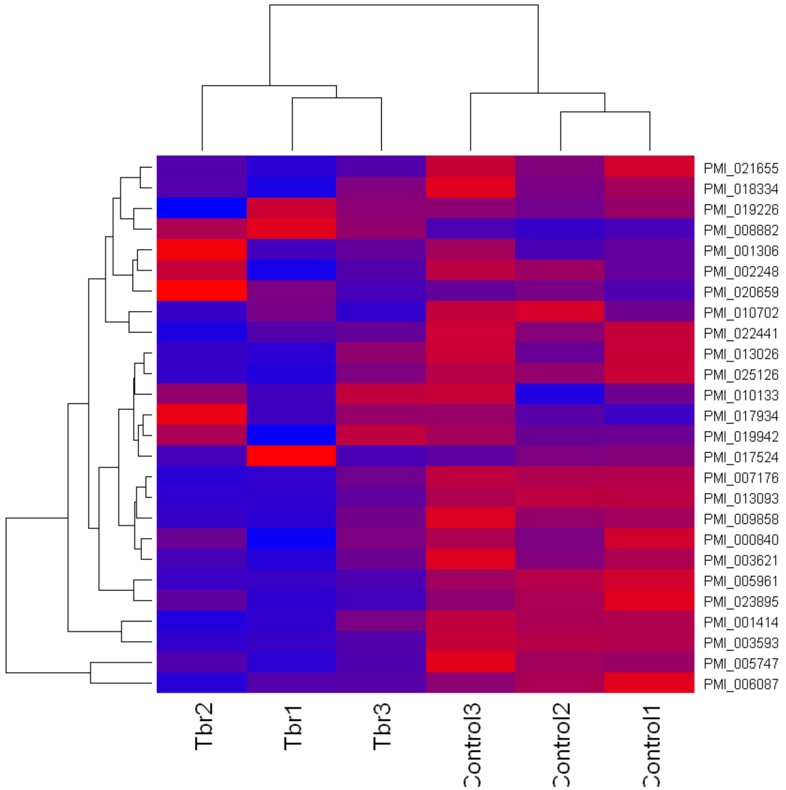
Clustering RNA-seq Data for an Application in Developmental Biology

4. We can explore the genes found in a reasonably small cluster.

In [28]:
```
# Based on your answers in the previous question, pick a cluster with a fairly small number of genes (say 50).
# The example below uses cluster 1, but the order of clusters is arbitrary so the best cluster to choose may be
# different for you.  Replace the cluster number below with your chosen cluster:

heatmap(tbrDataMatrix[kmeans4$cluster==1,],labRow=rownames(tbrDataGenes[kmeans4$cluster==1,]),col=palette)

# How would you characterize the expression pattern of these genes with respect to conditions and biological
# replicates?
```



Clustering Electrophysiology Traces for an Application in Neuroscience

C. Clustering analysis

1. Our analysis is going to be based on the hypothesis that there is some regularity to the data and that, specifically, neurons will partition into subgroups that show meaningfully different behaviors. We have already used one tool for identifying this kind of subgrouping: clustering. Before we get to more sophisticated analysis, we will see if clustering reveals any patterns among the neurons on dataset 1:

In [28]:
```
cluster1 <- kmeans(spikeCount1,3)
```

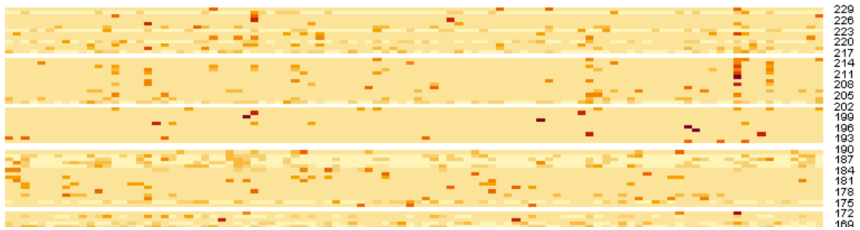
In [29]:
```
cluster1$cluster
```

```
1 3 1 2 1 1 2 1 1 1 1 1 1 1 1 3 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1
1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 3 1 1 2 1 1 1 1 1
1 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 2 1 1 1 1 1 3 2 1 2 1 2 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2
```

Do we see some clustering of neurons?

2. Let us try to visualize the distinction between the clusters based on the first 10s of data:

In [30]:
```
heatmap(spikeCount1[cluster1$cluster == 1,1:100],Rowv=NA,Colv=NA,labCol=NA)
```

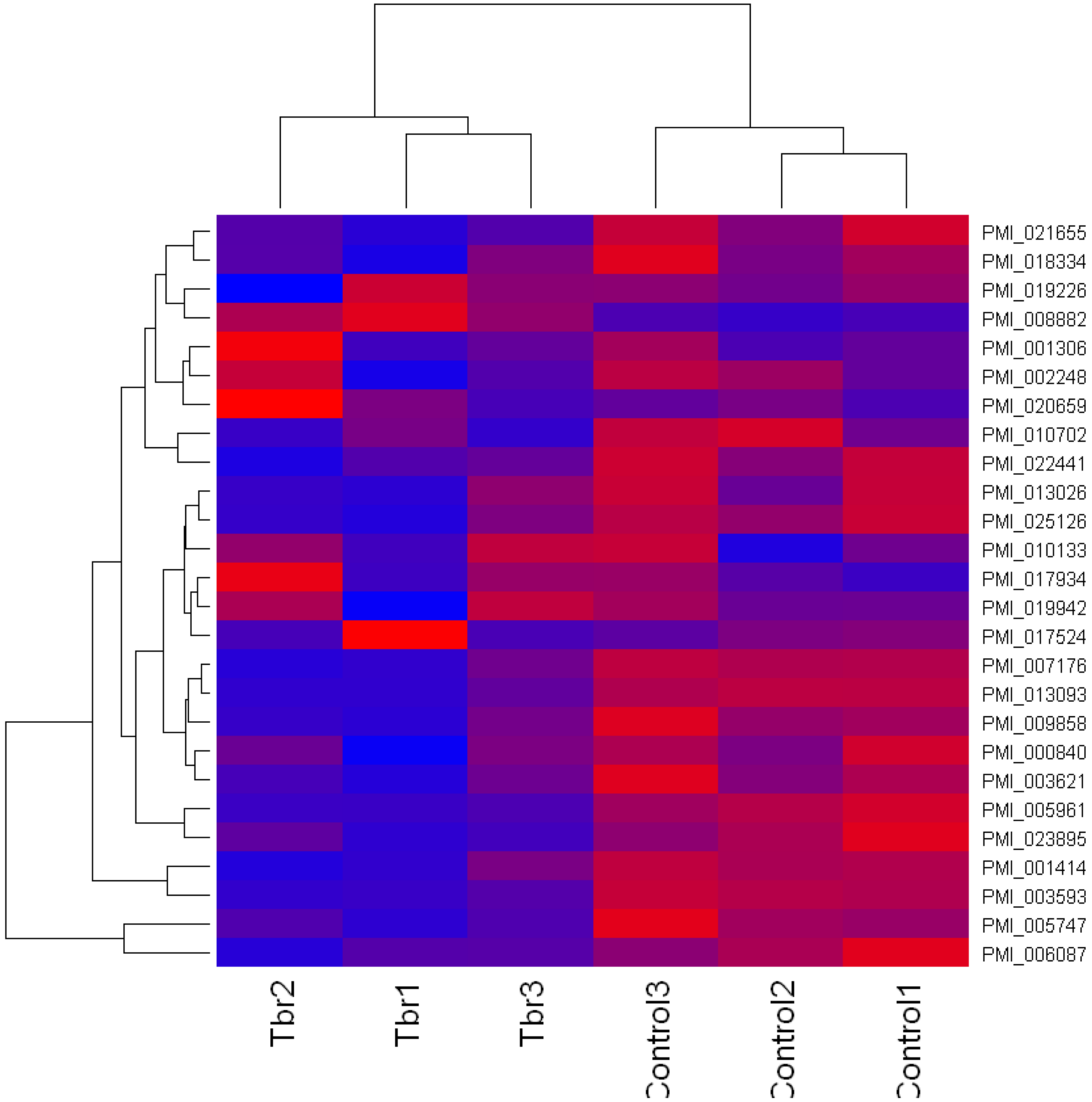


4. We can explore the genes found in a reasonably small cluster.

In [28]:
```
# Based on your answers in the previous question, pick a cluster with a fairly small number of genes (say 50).
# The example below uses cluster 1, but the order of clusters is arbitrary so the best cluster to choose may be
# different for you.  Replace the cluster number below with your chosen cluster:

heatmap(tbrDataMatrix[kmeans4$cluster==1,],labRow=rownames(tbrDataGenes[kmeans4$cluster==1,]),col=palette)

# How would you characterize the expression pattern of these genes with respect to conditions and biological
# replicates?
```



In [

ransfer across

stering Electrophysiology Traces
an Application in Neuroscience

g analysis

is going to be based on the hypothesis that there is some regularity to the data and that, urons will partition into subgroups that show meaningfully different behaviors. We have already or identifying this kind of subgrouping: clustering. Before we get to more sophisticated analysis, lustering reveals any patterns among the neurons on dataset 1:

```
means(spikeCount1,3)
```

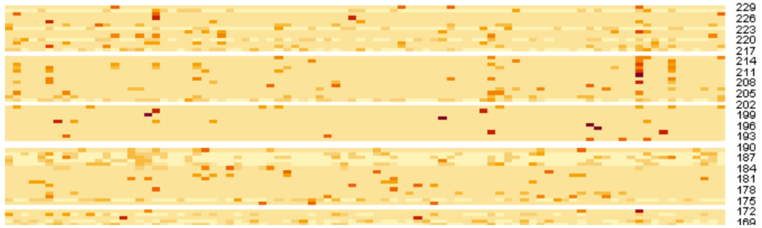
```
ster
```

```
1 2 1 1 1 1 1 1 1 1 3 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1
1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 3 1 1 2 1 1 1 1 1
1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 2 1 1 1 1 1 3 2 1 2 1 2 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 2 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2
```

e clustering of neurons?

visualize the distinction between the clusters based on the first 10s of data:

```
Count1[cluster1$cluster == 1,1:100],Rowv=NA,Colv=NA,labCol=NA)
```



C

d

Cluste

Applic

4. We can explo

```
In [28]:  # Based on yo
          # The example
          # different f

          heatmap(tbrDa

          # How would y
          # replicates?
```

C. Clustering analysis

1. Our analysis is going to be based on the hypothesis that there is some regularity to the data and that, specifically, neurons will partition into subgroups that show meaningfully different behaviors. We have already used one tool for identifying this kind of subgrouping: clustering. Before we get to more sophisticated analysis, we will see if clustering reveals any patterns among the neurons on dataset 1:

```
In [28]:  cluster1 <- kmeans(spikeCount1,3)
```

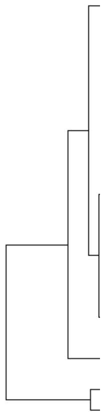
```
In [29]:  cluster1$cluster
```

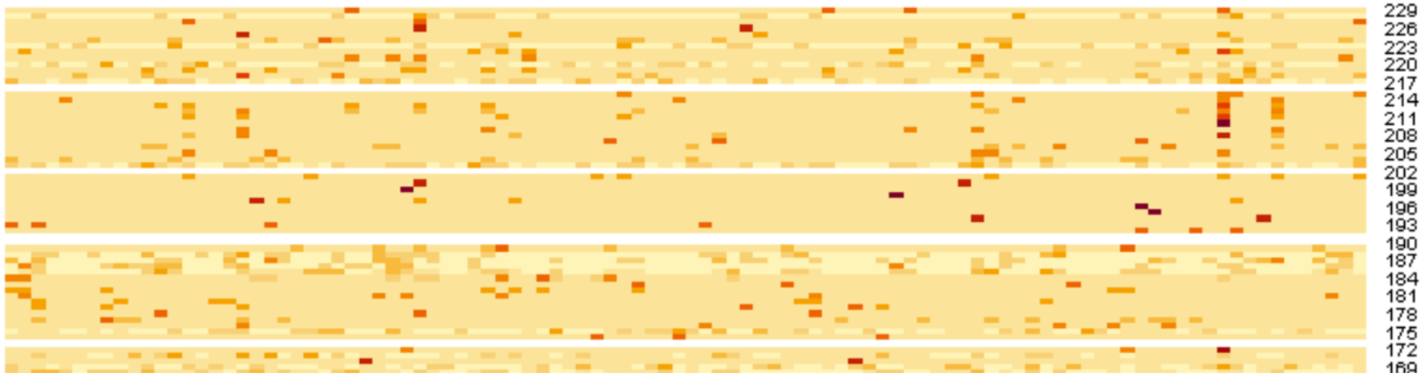
```
1 3 1 2 1 1 2 1 1 1 1 1 1 1 1 3 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1
1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 3 1 1 2 1 1 1 1 1
1 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 2 1 1 1 1 1 3 2 1 2 1 2 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2
```

Do we see some clustering of neurons?

2. Let us try to visualize the distinction between the clusters based on the first 10s of data:

```
In [30]:  heatmap(spikeCount1[cluster1$cluster == 1,1:100],Rowv=NA,Colv=NA,labCol=NA)
```





hat,
ve already
ed analysis,

1 1 1 2 1
1 1 1 1 1
1 1 1 1 1
1 1 1 1 1
1 2 1 1 1

“Workshops” using Jupyter Notebooks provide hands-on experience with real biological data

In [ ]:
```
# You can use the summary command to examine the transformed data.

summary(tbrDataLogNorm)
```

5. Try a t-test on a gene of interest to see if it shows differential expression:

In [ ]:
```
# We can first look at the data from a particular row of the matrix:
tbrDataLogNorm["PMI_005206",]

# Then run the actual t-test to see if it shows differential expression
t.test(x=tbrDataLogNorm["PMI_005206",c("Control1","Control2","Control3")],
       y=tbrDataLogNorm["PMI_005206",c("Tbr1","Tbr2","Tbr3")],var.equal=TRUE)$p.value

# You can try leaving off the `$p.value' from the end if you want to see more information than just the p-value
```

6. Try a t-test on every gene

In [ ]:
```
# The code below extends the test above to every gene in the table.
# This step will take a bit of time to run.

pvalue <- 0.01
significantGenes <- c()
for (i in 1:dim(tbrData)[1])
{
 if (sum(tbrData[i,c("Control1","Control2","Control3")])>10 &&
    sum(tbrData[i,c("Tbr1","Tbr2","Tbr3")])>10)
 {
  if (t.test(x=tbrDataLogNorm[i,c("Control1","Control2","Control3")],
             y=tbrDataLogNorm[i,c("Tbr1","Tbr2","Tbr3")],
             var.equal=TRUE)$p.value < pvalue)
  {significantGenes <-  append(significantGenes,rownames(tbrDataLogNorm[i,]))}
 }
}
```

7. Let’s look at some of the results:

In [ ]:
```
# See how many significant results were returned
```

“Workshops” using Jupyter Notebooks provide hands-on experience with real biological data

```
In [ ]: # You ca
        summary(

5. Try a

In [ ]: # We car
        tbrDataL

        # Then r
        t.test(x
               y

        # You ca

6. Try a

In [ ]: # The co
        # This s

        pvalue <
        signific
        for (i i
        {
         if (sum
            sum(
         {
          if (t.

          {signi
         }
        }

7. Let's

In [ ]: # See how many significant results were returned
```

5. Try a t-test on a gene of interest to see if it shows differential expression:

In [18]:

```
# We can first look at the data from a particular row of the matrix:
tbrDataLogNorm["PMI_005206",]

# Then run the actual t-test to see if it shows differential expression
t.test(x=tbrDataLogNorm["PMI_005206",c("Control1","Control2","Control3")],
       y=tbrDataLogNorm["PMI_005206",c("Tbr1","Tbr2","Tbr3")],var.equal=TRUE)

# You can try leaving off the `$p.value' from the end if you want to see more information than just the p-value
```

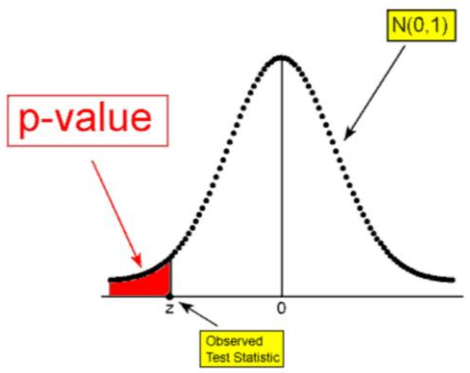
	Control1	Control2	Control3	Tbr1	Tbr2	Tbr3
PMI_005206	0.273329	0.353672	0.3070154	-0.9159592	-0.0135301	-0.6629596

```
        Two Sample t-test

data:  tbrDataLogNorm["PMI_005206", c("Control1", "Control2", "Control3")] and tbrDataLogNorm["PMI_005206", c("Tbr1", "Tbr2
", "Tbr3")]
t = 3.1218, df = 4, p-value = 0.03546
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.09316866 1.59114150
sample estimates:
 mean of x  mean of y
 0.3113388 -0.5308163
```

Interactive Learning w/ OLI Allows Self-Paced Introduction of Background Knowledge

Didactic material



Looking at the shaded region, you can see why this is often referred to as a **left-tailed** test. We shaded to the left of the test statistic, since less than is to the left.

Greater Than

The probability of observing a test statistic as **large as that observed or larger**, assuming that the values of the test statistic follow a standard normal distribution. Again, we will represent this probability in symbols and using the normal distribution.

Interactive exercises

If we are testing an alternative hypothesis of H_a: $p \neq p_0$, which of the following test statistics will give the smallest p-value?

z = hint −2

Correct. If z = −2, the data's p-hat is 2 standard deviations below p_0. So it is very unlikely that p-hats from random sampling will be located more standard deviations away from p_0 than the observed data. Hence the small p-value. Click here to see this illustrated using the distribution of z-scores.

Reset this Activity

Let's return to the scenario where we are studying the population of part-time college students. We know that in 2008, 60% of this population was female. We are curious if the proportion has decreased this year. We test the hypotheses: H_0: p = 0.60 and H_a: p < 0.60, where p is the proportion of part-time college students that are female this year.

Which of the following p-hat values will give the smallest p-value?

p-hat = hint --

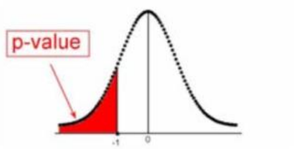
For the alternative hypothesis H_a: p < 0.60, the p-value is the area to the left of the test statistic.

Integrated assessments

Did I Get This?

In each of the following questions, choose the pair(s) of hypotheses for the population proportion (p) and the z statistic that match the figure.

Question 1:



- H_0: p = 0.56, H_a: p > 0.56, z = 1
- H_0: p = 0.56, H_a: p < 0.56, z = 1
- H_0: p = 0.56, H_a: p > 0.56, z = −1
- H_0: p = 0.56, H_a: p < 0.56, z = −1
- H_0: p = 0.56, H_a: p ≠ 0.56, z = 1 or H_0: p = 0.56, H_a: p ≠ 0.56, z = −1

Correct.

Images from the OLI Statistical Reasoning course – Module 13: Hypothesis testing for the population proportion p

Next Step for the Community: ISCB Course Affiliation/ Degree Endorsement

- We now have a much clearer idea of teaching practice in the field
  - ranges of personas and competencies they need in our field
  - balancing generality and specificity in making competencies usable
  - many examples of using them in teaching design, development, and assessment

- At the same time, this is a growing interest among practitioners in receiving guidance and some “stamp of approval” that what they are doing is sound

- As a result, ISCB has developed two evaluation mechanisms:
  - <u>Course affiliation</u>: a process to have short courses (hours to days of instruction on a single topic) assessed and approved by and then affiliated with the ISCB
  - <u>Degree endorsement</u>: a process to by which people offering degrees in computational biology or bioinformatics can have these certified by the ISCB

Mapping Objectives to Learning Experiences

Competencies

Learning Objectives

1. To be able to apply foundational mathematics and statistics needed for applied bioinformatics.
2. To have the breadth in basic biology to develop and apply bioinformatics methods to novel data sets.
3. To be able to independently undertake continued professional development in computational biology.

.
.
.

A. Work at depth in at least one technical area aligned with the life sciences.
B. Prepare life science data for computational analysis
C. Have a positive impact on scientific discovery through bioinformatics
D. Use data science methods suitable for the size and complexity of the data
E. Manage own and others' data according to community standards and principles
F. Make appropriate use of bioinformatics tools and resources
G. Contribute effectively to the design and development of user-centric bioinformatics tools and resources
H. Make appropriate and efficient use of scripting and programming languages
I. Construct, manage and maintain bioinformatics computing infrastructure of varying complexity
J. Comply with professional, ethical, legal and social standards and codes of conduct relevant to computational biology
K. Communicate meaningfully with a range of audiences - within and beyond your profession
L. Work effectively in teams to accomplish a common goal
M. Engage in continuing professional development in bioinformatics

Degree Program

Programming
Algorithms
Statistics
Cell Biology
Molecular Biology
Genomics Biology
Electives
CS Electives
Seminar Series
Capstone Project

.
.
.

Mapping Objectives to Learning Experiences

Competencies

Learning Objectives

1. To be able to apply foundational mathematics and statistics needed for applied bioinformatics.
2. To have the breadth in basic biology to develop and apply bioinformatics methods to novel data sets.
3. To be able to independently research and educate oneself in novel topics in nal biology.

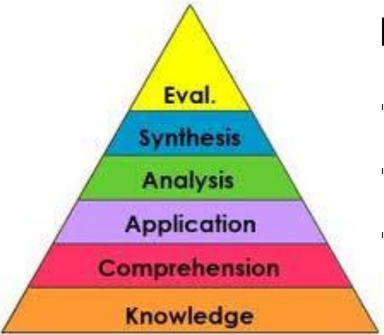


Image from Wikimedia Commons **File:Bloom taxonomy.jpg**, user Xristina la

A. Work at depth in at least one technical area aligned with the life sciences.
B. Prepare life science data for computational analysis
C. Have a positive impact on scientific discovery through bioinformatics
D. Use data science methods suitable for the size and complexity of the data
E. Manage own and others' data according to community standards and principles
F. Make appropriate use of bioinformatics tools and resources
G. Contribute effectively to the design and development of user-centric bioinformatics tools and resources
H. Make appropriate and efficient use of scripting and programming languages
I. Construct, manage and maintain bioinformatics computing infrastructure of varying complexity
J. Comply with professional, ethical, legal and social standards and codes of conduct relevant to computational biology
K. Communicate meaningfully with a range of audiences - within and beyond your profession
L. Work effectively in teams to accomplish a common goal
M. Engage in continuing professional development in bioinformatics

Degree Program

Programming
Algorithms
Statistics
Cell Biology
Molecular Biology
Genomics Biology
Electives
CS Electives
Seminar Series
Capstone Project
.
.
.

Mapping Objectives to Learning Experiences

Learning Objectives

1. To be able to apply foundational mathematics and statistics needed for applied bioinformatics.
2. To have the breadth in basic biology to develop and apply bioinformatics methods to novel data sets.
3. To be able to independently research and educate oneself in novel topics in nal biology.

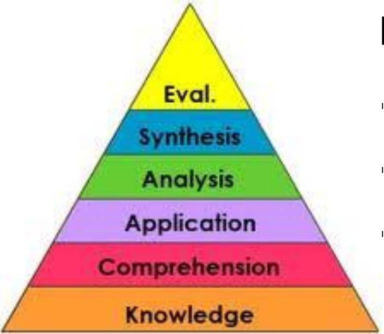


Image from Wikimedia Commons **File:Bloom taxonomy.jpg**, user Xristina la

Competencies

A. Work at depth in at least one technical area aligned with the life sciences.

B. Prepare life science data for computational analysis

C. Have a positive impact on scientific discovery through bioinformatics

D. Use data science methods suitable for the size and complexity of the data

E. Manage own and others' data according to community standards and principles

F. Make appropriate use of bioinformatics tools and resources

G. Contribute effectively to the design and development of user-centric bioinformatics tools and resources

H. Make appropriate and efficient use of scripting and programming languages

I. Construct, manage and maintain bioinformatics computing infrastructure of varying complexity

J. Comply with professional, ethical, legal and social standards and codes of conduct relevant to computational biology

K. Communicate meaningfully with a range of audiences - within and beyond your profession

L. Work effectively in teams to accomplish a common goal

M. Engage in continuing professional development in bioinformatics

Degree Program

Programming
Algorithms
Statistics
Cell Biology
Molecular Biology
Genomics
Biology Electives
CS Electives
Seminar Series
Capstone Project
.
.
.

Mapping Objectives to Learning Experiences

Learning Objectives

1. To be able to apply foundational mathematics and statistics needed for applied bioinformatics.
2. To have the breadth in basic biology to develop and apply bioinformatics methods to novel data sets.
3. To be able to independently research and educate oneself in novel topics in nal biology.

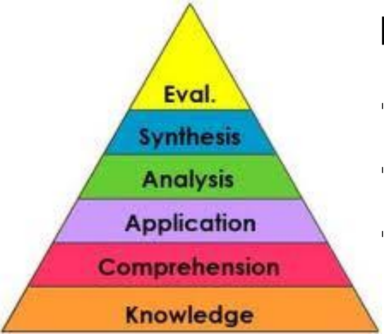


Image from Wikimedia Commons **File:Bloom taxonomy.jpg**, user Xristina la

Competencies

A. Work at depth in at least one technical area aligned with the life sciences.

B. Prepare life science data for computational analysis

C. Have a positive impact on scientific discovery through bioinformatics

D. Use data science methods suitable for the size and complexity of the data

E. Manage own and others' data according to community standards and principles

F. Make appropriate use of bioinformatics tools and resources

G. Contribute effectively to the design and development of user-centric bioinformatics tools and resources

H. Make appropriate and efficient use of scripting and programming languages

I. Construct, manage and maintain bioinformatics computing infrastructure of varying complexity

J. Comply with professional, ethical, legal and social standards and codes of conduct relevant to computational biology

K. Communicate meaningfully with a range of audiences - within and beyond your profession

L. Work effectively in teams to accomplish a common goal

M. Engage in continuing professional development in bioinformatics

Degree Program

Programming
Algorithms
Statistics
Cell Biology
Molecular Biology
Genomics
Biology Electives
CS Electives
Seminar Series
Capstone Project
.
.
.z

Mapping Your Degree's Profile to a Persona

A. Work at depth in at least one technical area aligned with the life sciences.

B. Prepare life science data for computational analysis

C. Have a positive impact on scientific discovery through bioinformatics

D. Use data science methods suitable for the size and complexity of the data

E. Manage own and others' data according to community standards and principles

F. Make appropriate use of bioinformatics tools and resources

G. Contribute effectively to the design and development of user-centric bioinformatics tools and resources

H. Make appropriate and efficient use of scripting and programming languages

I. Construct, manage and maintain bioinformatics computing infrastructure of varying complexity

J. Comply with professional, ethical, legal and social standards and codes of conduct relevant to computational biology

K. Communicate meaningfully with a range of audiences - within and beyond your profession

L. Work effectively in teams to accomplish a common goal

M. Engage in continuing professional development in bioinformatics

Existing Personas

Bioinformatics software engineer

Core facility director

Research group leader

Physician

Genetic counselor

Biocurator

Bioethicist

.

.

.

Novel Personas?

ISCB Web Application Form [Beta]

INTERNATIONAL SOCIETY FOR COMPUTATIONAL BIOLOGY

LEA... BIOLOGY AND BIOINFORMATICS
CONNECTING, TRAINING, EMPOWERING, WORLDWIDE

Password LOG IN
Forgot Login?

HOME | WHO WE ARE | WHAT WE DO | BECOME A MEMBER | CAREERS
search...

Short Course Affiliation and Degree-Program Endorsements Application

Please enter your Name and E-mail below, if editing a saved submission please also provide the Authorization Code:

Name: Russell Schwartz

Email: russells@andrew.cmu.edu

Authorization code: Only needed if editing a saved submission

Applying for: Degree Program Endorsement

Edit Saved Application

General information and Program Summary

Please enter the name, institution, and email of the principal applicant(s). All principal applicants must be members of ISCB.

Applicant Information

Name of principal applicant(s), one per line:

Russell Schwartz

Email address(es) of principal applicant(s), one per line:
NOTE: Please list emails in same order as applicants

russells@andrew.cmu.edu

Institutional affiliation:

Carnegie Mellon University

Official institution mailing address for correspondence:

5000 Fake Avenue
Faux F. Fictional Memorial Building
Pittsburgh, PA 15219

☑ I confirm that the principal applicants are members of ISCB at the time of application and for at least one year proceeding the application

Program Details

Name of the degree awarded:	BS in Computational Biology
Location(s) at which the degree program is offered: **NOTE: If offered at more than one campus, review must include all of the relevant performance sites**	Pittsburgh, PA
Language(s) in which program is taught:	English
Which agencies, aside from the offering university, if any, have certified the degree? (e.g., national degree certification agencies), one per line	N/A
How long has this degree program been in operation? **NOTE: The ISCB does not review degree programs until they have been in operation for at least three years**	20
How many cohorts have graduated to date? **NOTE: The ISCB does not review degree programs until they have graduated at least two cohorts**	16
Total number of program alumni:	75
Total number of students currently enrolled:	20
Minimum student effort required for the degree (in hours of time):	2400
URL to course website:	http://sample.compbio.cmu.edu

☐ Check this box if this is a request for a renewal of a previously ISCB-endorsed degree program

☐ Check this box if this is a request to reapply for a previously denied degree program

Program Plan

Program outline (<250 words): **Please comment on the purpose of your program, the applicant pool it is intended to serve, the current or intended size of the program, and summarize activities of the program to date.** **(196 words remaining)**	We offer a comprehensive program in general computational biology, covering background biological sciences and computer science training and specialized interdisciplinary training at the intersection. It is intended to prepare students for careers in computational biology or higher education in computational biology, related disciplines, or health progressions. The program is intended to accommodate
Program faculty, providing a list of course instructors, student advisors, program directors, and other key personnel, with titles and affiliations and role in the program. (We understand these roles may change from time to time, so please provide a current or recent snapshot.):	Course instructors: Russell Schwartz, Professor, CMU Random Q. Bioinformaticist, Asst Professor, CMU Alyssa X. Computer-Scientist, Professor, CMU Ben B. Biologist, Assoc. Professor, CMU Student advisors:
Summary of course requirements, listing required courses and restricted elective courses, total program requirements, and a sample curriculum	Required courses: Genetics Biochemistry Algorithms and Data Structures Machine Learning Computational Genomics Modeling and Simulation
Summary of mechanism for formative assessment to refine program design based on assessment results	Mechanisms for formative assessment: Student course evaluations Peer evaluation Department course reviews
How do you/will you ensure that the course is inclusive and non-discriminatory, consistent with the ISCB code of conduct(<250 words): **(199 words remaining)**	All courses are required to have a written plan for ensuring that content is inclusive and appropriately showcases roles of diverse scientists in the field, which are reviewed by the department head. University reporting mechanisms provide avenues for reporting of harassing or discriminatory conduct by instructors and students, including anonymous reporting and

Short Course Affiliation and Degree-Program Endorsements Application

Please enter the requirements for this degree program (I.E. the courses that make up the program):
NOTE: One per line

Algorithms and Data Structures
Computational Genomics
Modeling and Simulation
Advanced Biology Elective
Advanced Computer Science Elective
Advanced Computational Biology Elective
Professional Development

List the expected outcomes for this degree program:
NOTE: One per line

of computational models.
Ability to select appropriate algorithms and data structures for a defined task.
Ability to employ appropriate statistical tools for biological data analysis
Ability to work on an interdisciplinary team in applying computational tools to biomedical problems.

Save Application | Map Required Competencies

Mapping degree requirements to competencies

Please apply the bloom taxonomy to to each idenitified **requirement** to the ISCB Competencies. If you see an error in your requ

Please apply the bloom taxonomy to to each idenitified **requirement** to the ISCB Competencies. If you see an error in your requ

	A3: Work at depth in at least one technical area aligned with the life sciences	B3: Prepare life science data for computational analysis	C3: Have a positive impact on scientific discovery through bioinformatics	D3: Use data science methods suitable for the size and complexity of the data	E3: M c a comm ar
Biochemistry	Analysis	None	None	None	None
Genetics	Analysis	Knowledge	None	None	-- Sel
Cell Biology	Analysis	None	None	None	-- Sel
Programming	None	None	None	Application	-- Sel
Algorithms and Data Structures	None	None	None	Analysis	-- Sel
Computational Genomics	Application	Comprehension	Application	Application	-- Sel
Modeling and Simulation	Application	Comprehension	Application	Application	-- Sel
Advanced Biology Elective	Evaluation	None	None	None	-- Sel
Advanced Computer Science Elective	None	None	None	Evaluation	-- Sel
Advanced Computational Biology Elective	Synthesis	Synthesis	Application	Evaluation	-- Sel
Professional Development	None	None	None	None	-- Sel

Mapping program outcomes to competencies

Please apply the bloom taxonomy to to each idenitified **outcome** to the ISCB Competencies. If you see an error in your outcomes please us

	A3: Work at depth in at least one technical area aligned with the life sciences	B3: Prepare life science data for computational analysis	C3: Have a positive impact on scientific discovery through bioinformatics	D3: Use data science methods suitable for the size and complexity of the data	E3: Manage ow others' da according community sta and princip
bility to apply eneral biology nowledge in evelopment of omputational nodels.	Analysis	Application	Application	Analysis	Knowledge
bility to select ppropriate lgorithms and ata structures or a defined ask.	None	Application	Application	Evaluation	None
bility to mploy ppropriate tatistical tools or biological ata analysis	Application	Analysis	Application	Application	Knowledge
bility to work n an nterdisciplinary eam in pplying omputational ools to iomedical roblems.	Application	Application	Application	Application	Application

Mapping degree requirements to co

Mapping program outcomes to

Please apply t

Please app

Biochem
Genetics
Cell Biolo
Program
Algorithr and Data Structure
Computa Genomic
Modeling Simulatio
Advance Biology Elective
Advance Compute Science Elective
Advance Computa Biology Elective
Professi Develop

	A3: Work at depth in at least one technical area aligned with the life sciences	B3: Prepare life science data for computational analysis	C3: Have a positive impact on scientific discovery through bioinformatics	D3: Use methods the comple
Biochemistry	Analysis	None	None	None
Genetics	Analysis	Knowledge	None	None
Cell Biology	Analysis	None	None	None
Programming	None	None	None	Applicat
Algorithms and Data Structures	None	None	None	Analysis
Computational Genomics	Application	Comprehension	Application	Applicat
Modeling and Simulation	Application	Comprehension	Application	Applicat

comes please us

E3: Manage ow others' da according community sta and princip

Knowledge

None

Knowledge

Application

The Degree Program Review Process [Draft]

- **Phase 0**: Applicants submit an initial application for prescreening
  - Application describes program and provides evidence of credibility, rigor
  - Nominal application fee to discourage frivolous applications
  - ISCB staff provide a quick screen for completeness and relevance
- **Phase 1**: Credible applications go out for a full review by a journal-like review committee
  - May be multiple rounds of review/revision
  - End result is rejection or Phase 2 review
- **Phase 2**: Due diligence by ISCB
  - Options for certifying program is real and facilities exist to offer it (site visit, virtual site visit, endorsement by trusted endorsers)
  - End result is rejection, acceptance, or remediation plan
- **Endorsement**: program is endorsed by the ISCB for 5 years (with some caveats)
  - Program gets acknowledgement from ISCB and use of ISCB endorsement logo
  - Program gets some ISCB advertisement (newsletter/website)

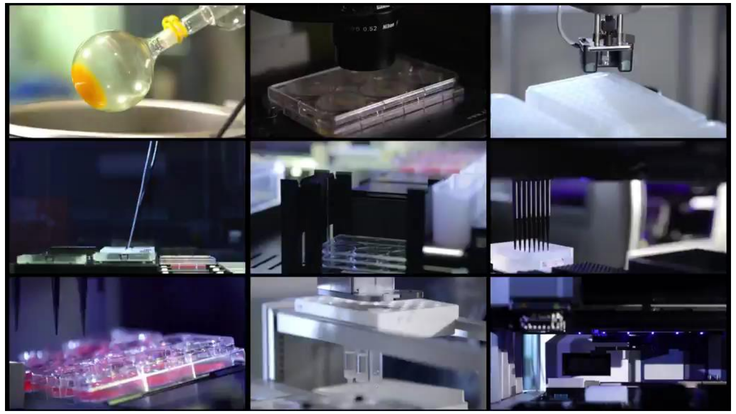
CMU Perspective on the Future of Computational Education for the Life Sciences: Automating Biological Research

- Incubation for cell culture
- Liquid handling
- Microscopy
- Nucleic Acid extraction
- Q-PCR
- Centrifugation
- Flow Cytometry/ Cell Sorter

The MSAS Philosophy:
Advanced automation requires two kinds of systems

1. Systems for performing repetitive tasks
    - Ex. Milking Cows; Grinding & Welding; Liquid handling





These systems operate in highly controlled environments
For the most part, these machines do not need to solve problems (i.e., handle novel situations)

The MSAS Philosophy:
Advanced automation requires two kinds of systems

2. Systems that can 'think' and handle novel challenges
   - This is an emerging area of automation that leverages advances in computing power, artificial intelligence, and machine learning

- Examples

IBM Watson

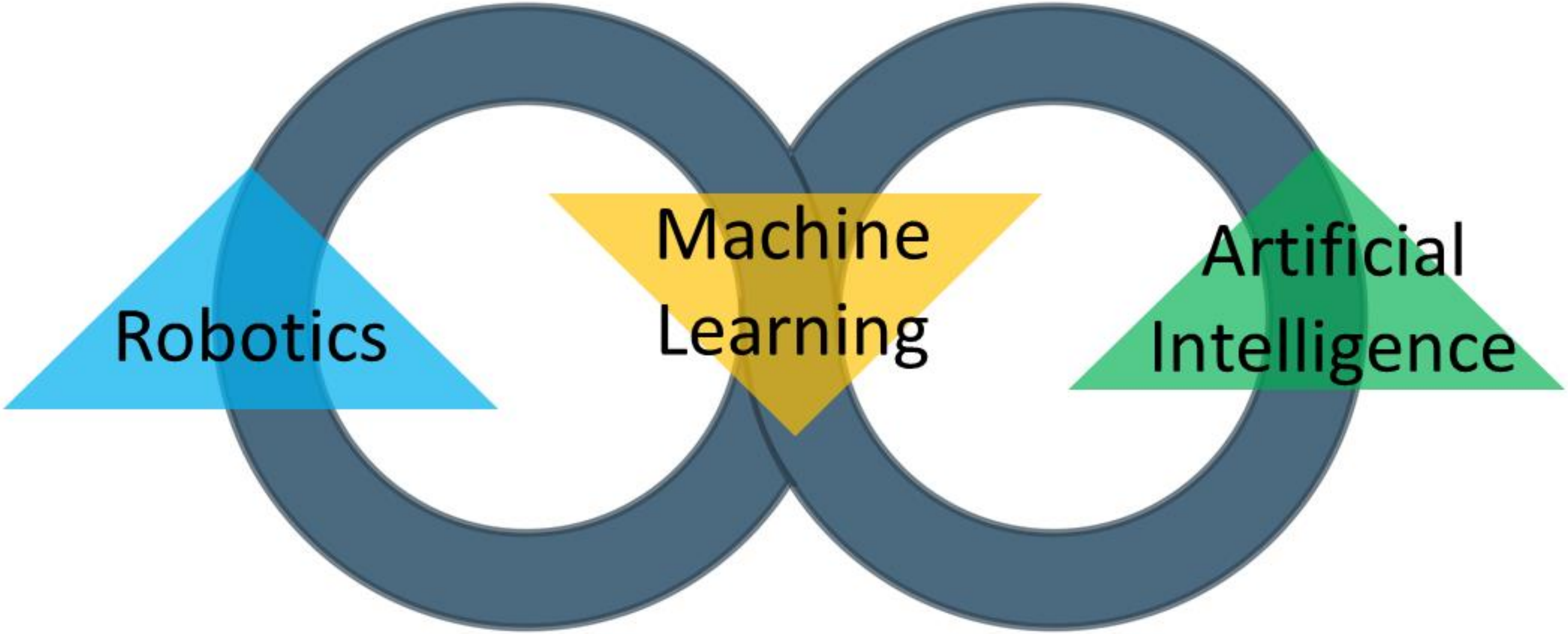
NewsBot


Amazon Robots

Self-driving Cars

These systems operate in largely **uncontrolled environments**
They **do** encounter novel situations & solve problems

Next Step for Us: Automating Scientific Research (MS in Automated Science)



Automatic
Data Generation

Automatic
Data Interpretation

Automatic
Experiment selection

Automated Lab Methods
02-761 & 02-762 (Year 1)

Machine Learning
for Scientists
02-620 (2nd semester)

Automation of Scientific Research
02-750 (2nd semester)

Democratizing science education via automated and online instruction

Building a Research-Based Bioinformatics Education Program for High School Students

Josh Kangas, Ph.D.

Assistant Teaching Professor

Phillip Compeau, Ph.D.

Associate Teaching Professor

Conclusions

- There is now a large community of bioinformatics educators who can benefit from one another's experience about what to teach and how to teach it
- A common analysis framework --- competencies/personas --- can help us compare experiences, although finding the right resolution is a challenge
- When we can recognize commonalities among our training experiences, we can find ways to serve many populations with overlapping needs
- This is an ongoing process of each of us developing ideas in our home institutions and working with colleagues internationally to share experiences and learn from one another

Acknowledgments

Introduction to Computational Biology

- Robert Murphy
- Elizabeth Jones

Data Science for Biologists

- Christina Akirtava (TA)
- Co-instructors:
  - Veronica Hinman
  - Sandra Kuhlman
  - Luisa Hiller
  - Alison Barth
- Aaron Mitchell
- Laura Ochs Pottmeyer
- Emily Daniels Weiss
- The Eberly Center for Teaching Excellence & Educational Innovation

Automated Science

- Josh Kangas
- Robert Murphy
- Phillip Compeau

The ISCB Curriculum Task Force and Affiliates

- Nicola Mulder, Computational Biology group, Department of Integrative Biomedical Sciences, IDM, University of Cape Town, Cape Town, South Africa
- Cath Brooksbank, EMBL-EBI, Wellcome Genome Campus, Hinxton, Cambridge, UK
- Michelle D. Brazas, Ontario Institute for Cancer Research, Toronto, Canada
- Sarah Morgan, EMBL-EBI, Wellcome Genome Campus, Hinxton, Cambridge, UK
- Gabriella Rustici, University of Cambridge, Cambridge, UK
- Lonnie Welch, School of Electrical Engineering and Computer Science, Ohio University, Athens, Ohio, USA
- Mark Pauley, School of Interdisciplinary Informatics, University of Nebraska at Omaha, Omaha, Nebraska, USA
- Anne Rosenwald, Department of Biology, Georgetown University, Washington, D.C., USA
- Tandy Warnow, Department of Computer Science, University of Illinois, Urbana-Champaign, USA
- Bruno Gaeta, School of Computer Science and Engineering, UNSW Australia, Sydney, Australia
- Alastair M. Kilpatrick, Department of Pediatrics, University of California San Diego, La Jolla, CA, USA
- Daniel Mietchen, National Institutes of Health, Bethesda, MD USA
- Benjamin L. Moore, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK
- William Pearson, Department of Biochemistry and Molecular Genetics, School of Medicine, University of Virginia, Charlottesville, VA USA
- Predrag Radivojac, School of Informatics and Computing, Indiana University, Bloomington, IN, USA
- Naomi Rosenberg, Tufts University School of Medicine, Boston, MA, USA
- Michael Sierk, St. Vincent College, Latrobe, PA USA
- workshop partiticpants too numerous to list

Open Learning Initiative
Transforming higher education through the science of learning.

EMBL-EBI iSCB INTERNATIONAL SOCIETY FOR COMPUTATIONAL BIOLOGY elixir CABANA Capacity building for bioinformatics in Latin America H3ABioNet