Biomedical Data Science Curricula at the University of Wisconsin

Mark Craven
Department of Biostatistics & Medical Informatics
Department of Computer Sciences
Relevant programs at the University of Wisconsin

- CIBM training grant (NLM-funded T15)
- BD2K training grant (BD2K/NLM-funded T32)
- MS program in Biomedical Data Science
- PhD program in Biomedical Data Science
- summer research program in Biomedical Data Science
Computation and Informatics in Medicine and Biology (CIBM) program

- trainees are recruited from a broad set of PhD programs/departments/centers including

<table>
<thead>
<tr>
<th>recent predocs</th>
<th>recent postdocs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biochemistry</td>
<td>Bacteriology</td>
</tr>
<tr>
<td>Biomedical Engineering</td>
<td>Biochemistry</td>
</tr>
<tr>
<td>Chemistry</td>
<td>Biostatistics &amp; Medical Informatics</td>
</tr>
<tr>
<td>Clinical Investigation</td>
<td>Genome Center of Wisconsin</td>
</tr>
<tr>
<td>Computer Sciences</td>
<td>Marshfield Clinic Research Institute</td>
</tr>
<tr>
<td>Epidemiology</td>
<td>Psychology</td>
</tr>
<tr>
<td>Genetics</td>
<td>Psychiatry</td>
</tr>
<tr>
<td></td>
<td>Statistics</td>
</tr>
<tr>
<td></td>
<td>School of Veterinary Medicine</td>
</tr>
<tr>
<td></td>
<td>Waisman Center (human development)</td>
</tr>
</tbody>
</table>
CIBM training approach

• CS prerequisites: *Intro to Programming, Data Structures*
• all trainees should gain solid grounding in both quantitative methods and biomedicine (but each trainee typically has an 80/20 mix of expertise)
• core courses ensure all trainees understand central problems and approaches in biomedical informatics
• dual mentorship
CIBM Curriculum

• core courses
  
  *Introduction to Bioinformatics*
  
  *Health Informatics*
  
  *Introduction to Biostatistics*

• $\geq 2$ courses in biomedical sciences
• $\geq 1$ advanced course in biomedical informatics
• $\geq 1$ advanced course in computer science
• course in *Responsible and Ethical Conduct of Research*
• CIBM seminar course every semester
Some course options

• advanced courses in biomedical informatics
  Medical Image Analysis
  Advanced Bioinformatics
  Modeling Biological Systems
  Decision Making in Health Care
  ...

• advanced courses in computer science
  Machine Learning
  Computer Vision
  Intro to Human-Computer Interaction
  Database Management Systems
  Linear Programming
  Introduction to Data Science
  ...

MS Program in Biomedical Data Science

- students come from a broad range of backgrounds: undergrad degrees in CS/engineering, PhD in biological sciences, PharmD, MD
- CS prerequisites: *Intro to Programming, Data Structures*
- all students should gain solid grounding in both data science methods and biomedicine
- core courses ensure all students understand central problems and approaches in biomedical informatics
- the courses for a student should have a focus in terms of
  - area of quantitative biomedical studies
  - data science methodology
MS in Biomedical Data Science Curriculum

• core courses
  
  *Introduction to Bioinformatics*
  
  *Health Informatics*
  
  *Medical Image Analysis*
  
  *Introduction to Biostatistics*

• 2 “concentration” electives

• 2 data science electives

• \( \approx 2 \) “track” electives
MS in Biomedical Data Science Curriculum

• concentration electives
  
  Medical Image Analysis
  Advanced Bioinformatics
  Modeling Biological Systems
  Decision Making in Health Care
  Statistical Methods for Clinical Trials
  Statistical Methods for Epidemiology
  Statistical Methods for Molecular Biology

…
MS in Biomedical Data Science Curriculum

- data science electives
  - Machine Learning
  - Computer Vision
  - Intro to Human-Computer Interaction
  - Database Management Systems
  - Linear Programming
  - Introduction to Data Science
  - Mathematical Statistics and Inference
  - Statistical Computing
  - Theory and Application of Regression
  ...

New course: Ethical Conduct of Research for Data Scientists

• being developed by Prof. Pilar Ossorio
• centered on 8 case studies that are built around the real-world experiences of biomedical data scientists
• materials to be made available

**Case 1:** Constructing Genomic Signatures of Oncology Treatment Response

**Focus topics:**
- Data integrity and reproducibility
- Research misconduct
- Mentor-mentee responsibilities and relationships

**Case 2:** Developing an Algorithm to Predict Breast Cancer Recurrence Risk using EHR and Genotype Data: Collaboration with a Health System Partner

**Focus topics:**
- Privacy, data anonymization, and re-identification risk for data subjects
- HIPAA privacy rule
- Federal policy for the protection of human research subjects: use of coded data from humans
- Scientist as a responsible member of society: conducting trustworthy science
New course: Data Analysis and Visualization

• being developed by Prof. Karl Broman
• to be held in conjunction with his *Tools for Reproducible Research* course
• materials to be made available
• key topics
  • managing and manipulating heterogeneous data files
  • data diagnostics and cleaning
  • data visualization
  • exploratory data analysis
  • formulating and identifying appropriate statistical models and methods
  • simulation-based methods