Biomedical Data Science at Vanderbilt

December 13, 2017

Bradley Malin, Ph.D.
Vice Chair for Research, Department of Biomedical Informatics
Professor of Biomedical Informatics, Biostatistics, & Computer Science
Vanderbilt University

Informatics Big Picture

- Department of Biomedical Informatics situated in the School of Medicine
- Over 40 primary faculty and 40 secondary faculty
- Approximately 40 trainees
  - MS / PhD
  - MSACI
  - Clinical Informatics Fellows
  - Undergraduate and high school interns
- NLM T15 (with data science supplement)
- NLM T32 (big data science)

T15 Principle Investigators

Bradley Malin, Ph.D.
Biomedical Informatics, Biostatistics & Computer Science

Cindy Gadd, Ph.D.
Biomedical Informatics

Gretchen Purcell Jacks, M.D., Ph.D.
Biomedical Informatics & Surgery

https://news.vanderbilt.edu/2016/04/21/new-doctoral-track-focuses-on-big-biomedical-data-science/

Vanderbilt BIDS: Big Biomedical Data Science Training Program

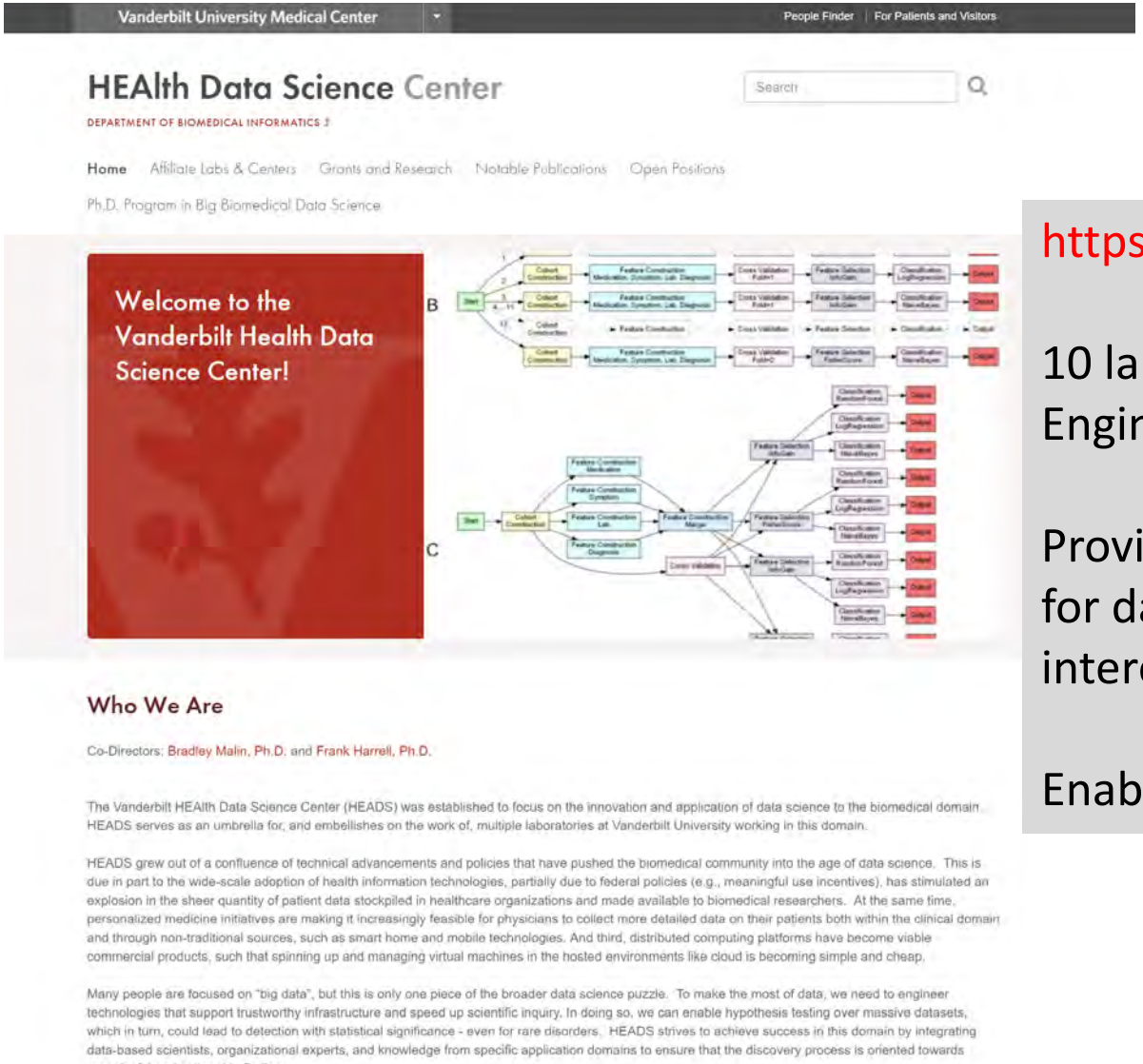
Cindy Gadd, Ph.D.
Biomedical Informatics

Jeffrey Blume, Ph.D.
Biostatistics

Bradley Malin, Ph.D.
Biomedical Informatics, Biostatistics & Computer Science

https://www.vumc.org/dbmi/vanderbilt-big-biomedical-data-science-bids-program

Health Data Science Center



https://www.vumc.org/heads/

10 labs in School of Medicine, Engineering, and Arts & Sciences

Provides guidance and support for data science activities and interests

Enables a home for T32 students

Primary BIDS Faculty Appointments

School of Medicine

- Anesthesiology
- Biochemistry
- Biomedical Informatics
- Biostatistics
- Cancer Biology
- Genetics
- Health Policy
- Medicine
- Molecular Physiology & Biophysics
- Hematology / Oncology

School of Engineering

- Biomedical Engineering
- Computer Engineering
- Computer Science
- Electrical Engineering

School of Arts & Sciences

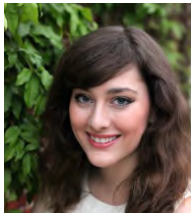
- Biology
- Chemistry

3 Formal Levels of Education

- Ph.D. in Biomedical Informatics: Data Science Track (DST)
  - 5 year commitment
- Ph.D. in Another Field: M.S. in DST
  - 2-3 year commitment
- Ph.D. in Another Field: Course & Research Sponsorship
  - 1 year renewable (but competitive)

2016 – 2017

- Ph.D. in Biomedical Informatics: Data Science Track (DST)



Mary Lauren Benton

Evolutionary Genomics

- Ph.D. in Another Field: M.S. in DST

Michael Pritchard
(Ph.D. Student in CS)

Adversarial Learning for (Anti-)Viral Systems

Nick Strayer
(Ph.D. Student in Biostats)

Visualization of Big Statistics

- Ph.D. in Another Field: Course & Research Sponsorship

Lucy D'Agustino
(Ph.D. Student in Biostats)

New Approaches to Significance Scoring

Andrew Plassard
(Ph.D. Student in CS)

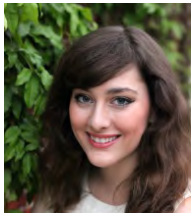
Big Brain Image Processing

Rohit Venkat
(Ph.D. Student in Molecular Physiology & Biophysics)

Big Math Models of Cellular Signaling

2017 – 2018

- Ph.D. in Biomedical Informatics: Data Science Track (DST)



Mary Lauren Benton

Evolutionary Genomics

Kim Kondratieff

Phenomics

Grayson Ruhl

Data Privacy

- Ph.D. in Another Field: M.S. in DST

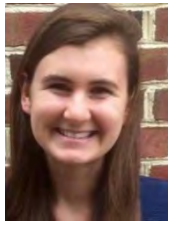
Michael Pritchard
(Ph.D. Student in CS)

Adversarial Learning for
(Anti-)Viral Systems

Nick Strayer
(Ph.D. Student in Biostats)

Visualization of Big Statistics

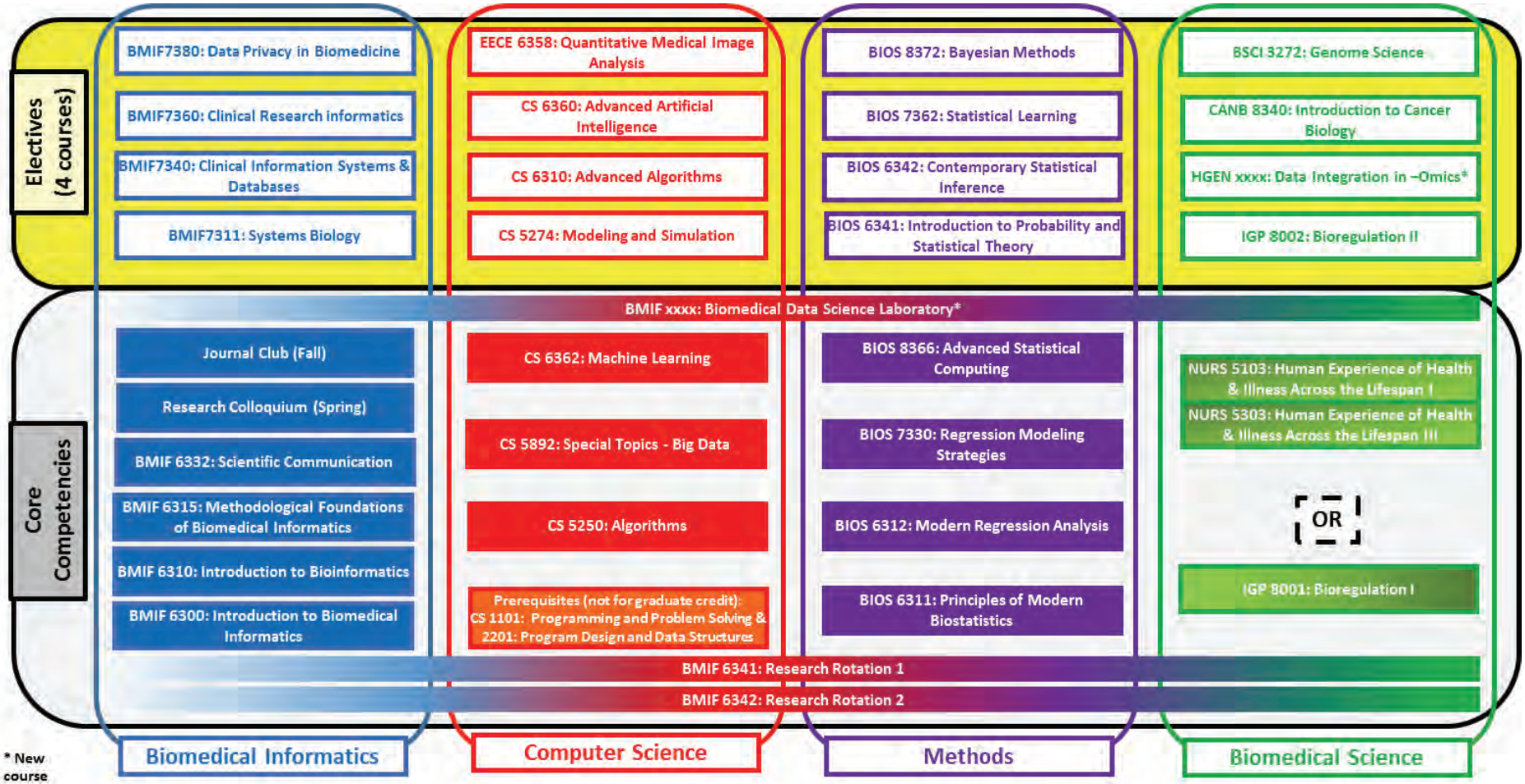
- Ph.D. in Another Field: Course & Research Sponsorship

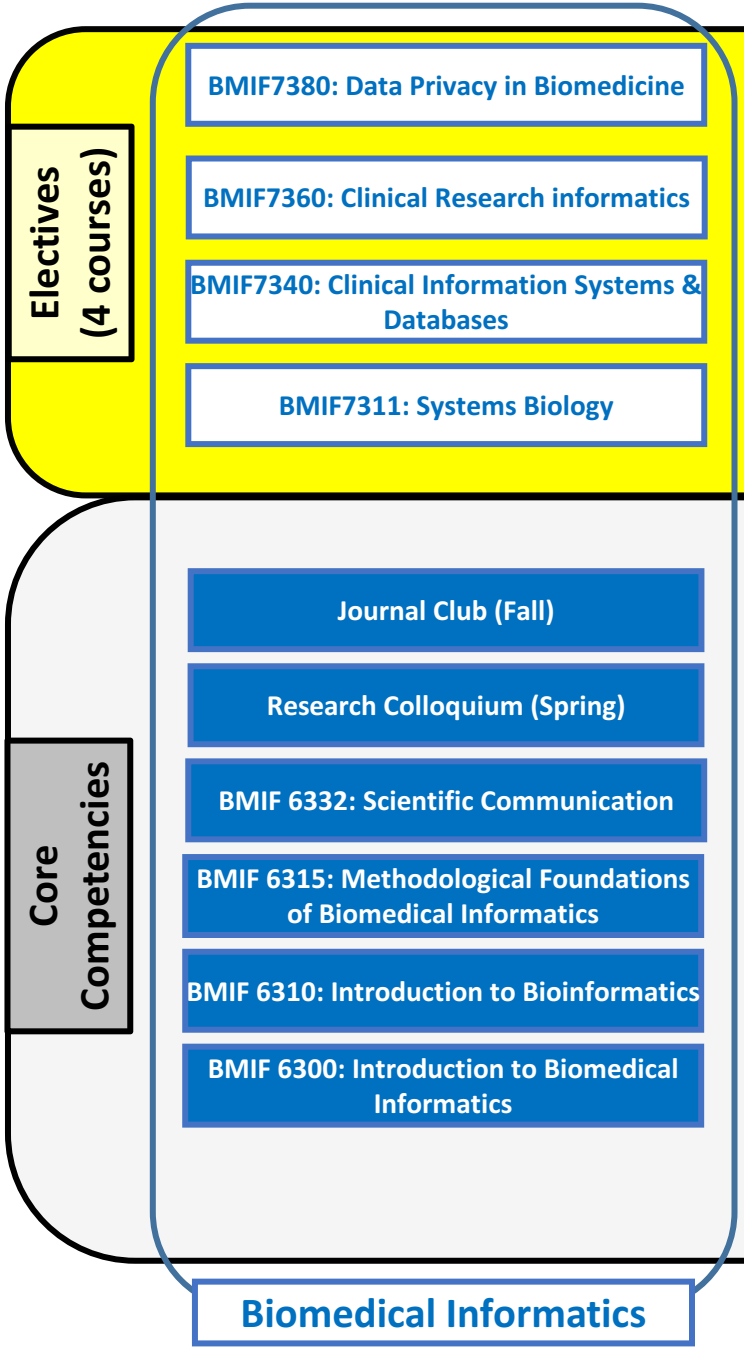


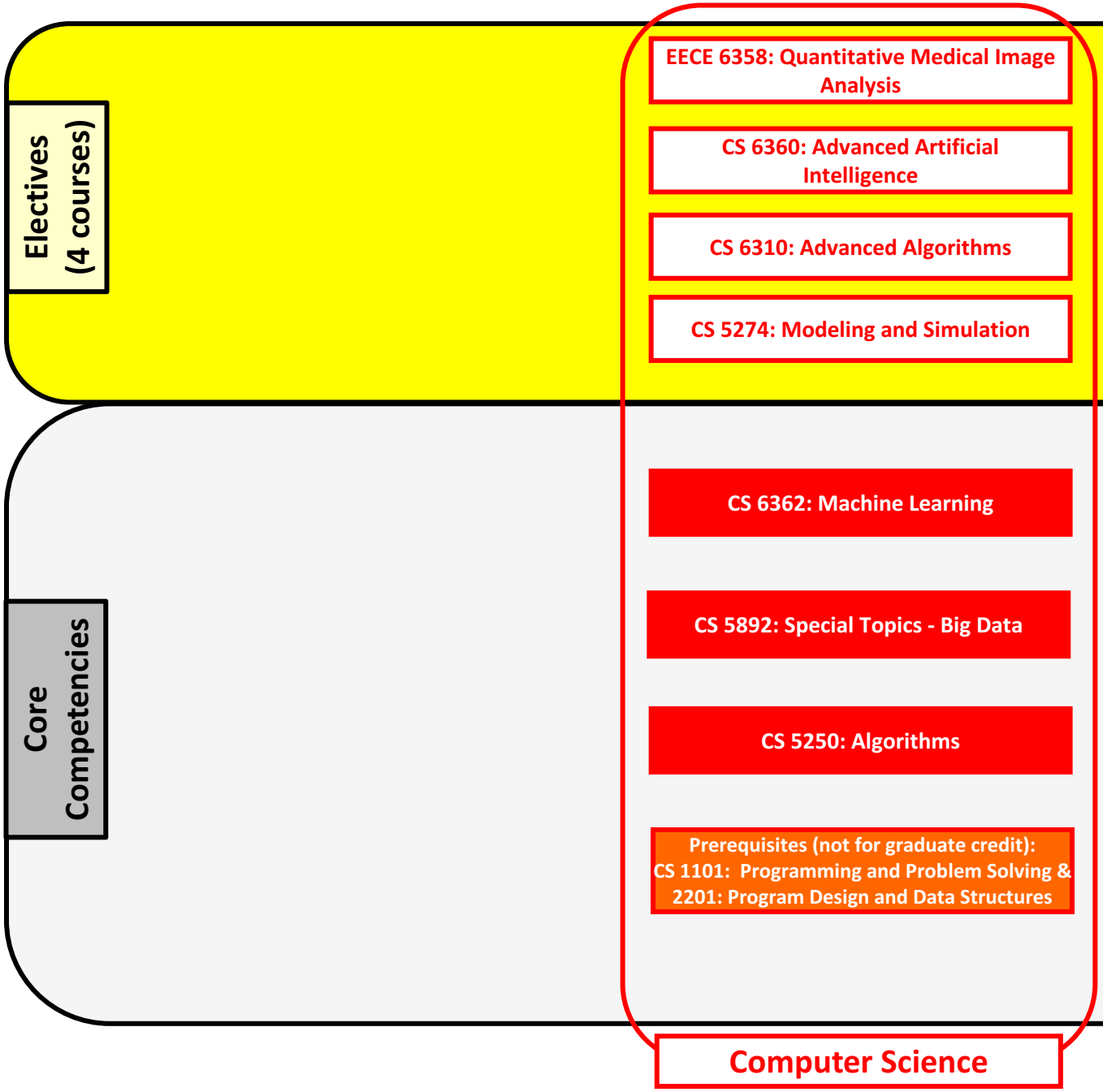
Sarah Maddox
(Ph.D. Student in Biochemistry)

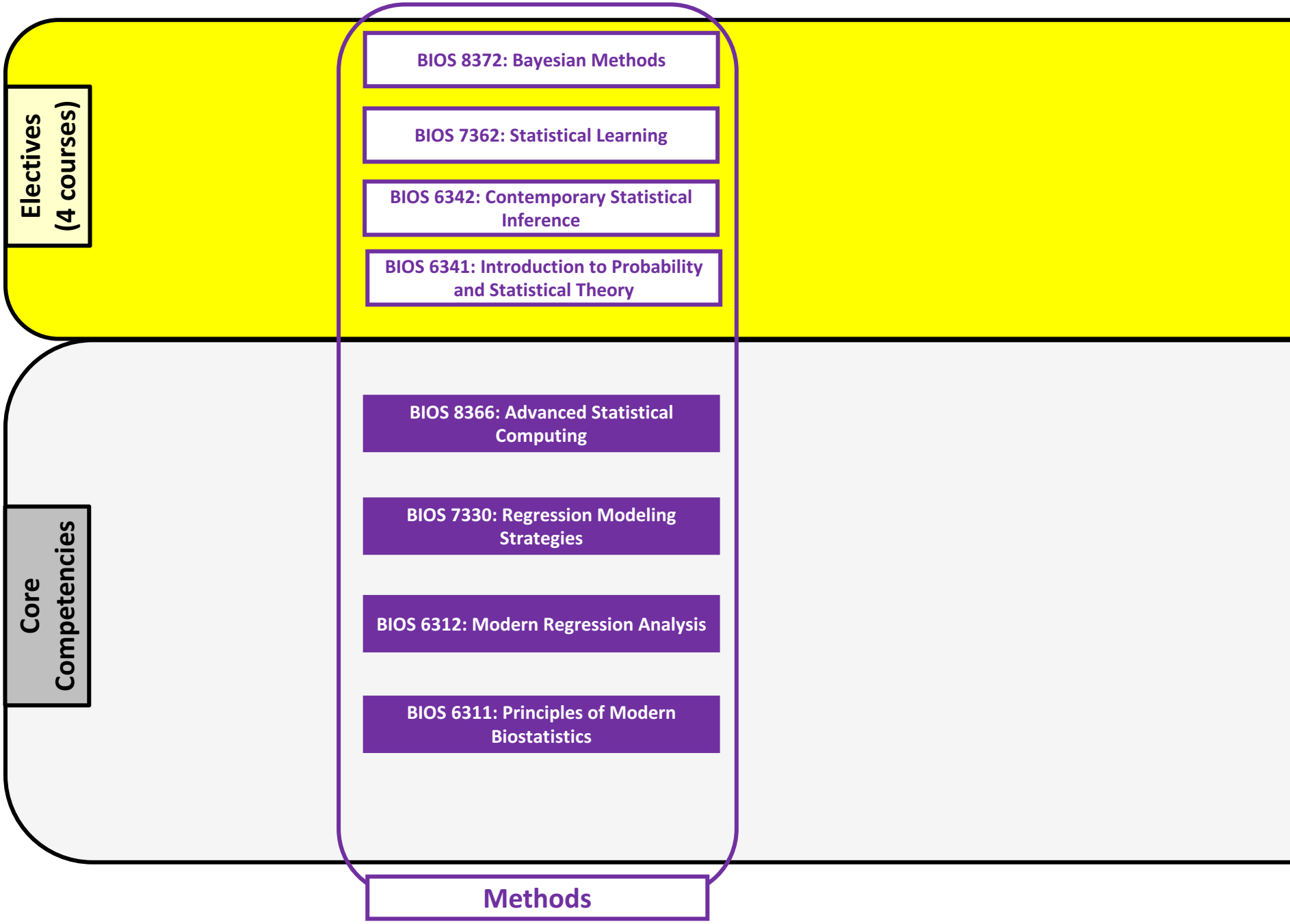
Phenotypic Heterogeneity in
Small Cell Lung Cancer

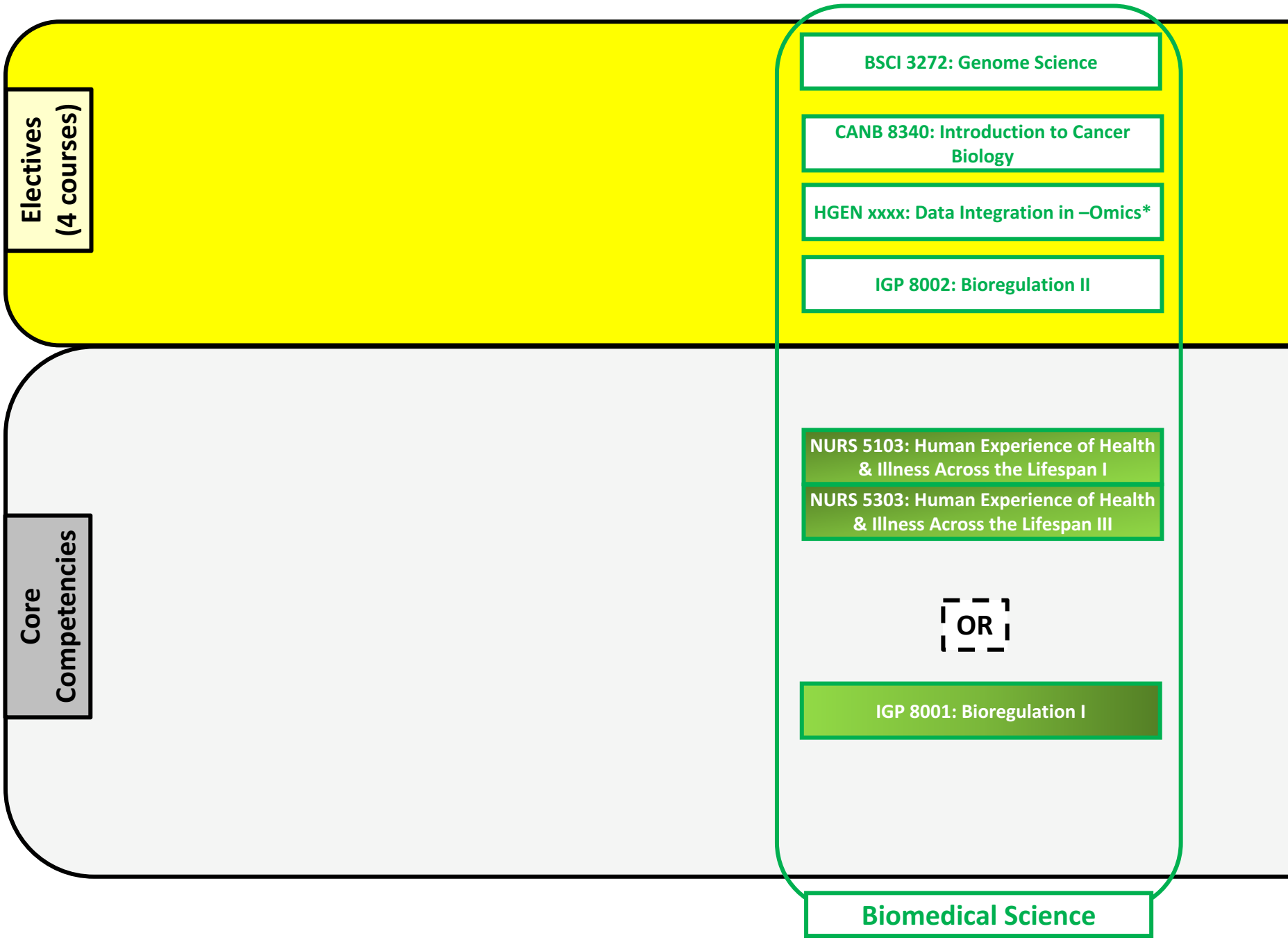
BIDS Curriculum




Electives (4 courses)
BMIF7380: Data Privacy in Biomedicine
BMIF7360: Clinical Research informatics
BMIF7340: Clinical Information Systems & Databases
BMIF7311: Systems Biology
Core Competencies
Journal Club (Fall)
Research Colloquium (Spring)
BMIF 6332: Scientific Communication
BMIF 6315: Methodological Foundations of Biomedical Informatics
BMIF 6310: Introduction to Bioinformatics
BMIF 6300: Introduction to Biomedical Informatics
Biomedical Informatics


Electives (4 courses)
EECE 6358: Quantitative Medical Image Analysis
CS 6360: Advanced Artificial Intelligence
CS 6310: Advanced Algorithms
CS 5274: Modeling and Simulation
Core Competencies
CS 6362: Machine Learning
CS 5892: Special Topics - Big Data
CS 5250: Algorithms
Prerequisites (not for graduate credit):
CS 1101: Programming and Problem Solving &
2201: Program Design and Data Structures
Computer Science


Electives (4 courses)
BIOS 8372: Bayesian Methods
BIOS 7362: Statistical Learning
BIOS 6342: Contemporary Statistical Inference
BIOS 6341: Introduction to Probability and Statistical Theory
Core Competencies
BIOS 8366: Advanced Statistical Computing
BIOS 7330: Regression Modeling Strategies
BIOS 6312: Modern Regression Analysis
BIOS 6311: Principles of Modern Biostatistics
Methods


Electives (4 courses)
Core Competencies
BSCI 3272: Genome Science
CANB 8340: Introduction to Cancer Biology
HGEN xxxx: Data Integration in –Omics*
IGP 8002: Bioregulation II
NURS 5103: Human Experience of Health & Illness Across the Lifespan I
NURS 5303: Human Experience of Health & Illness Across the Lifespan III
OR
IGP 8001: Bioregulation I
Biomedical Science

T15 Supplement – Data Privacy Course (http://hiplab.mc.vanderbilt.edu/courses/BMIF380/)

BMIF-7380 / CS-8396-02: Data Privacy in Biomedicine

Main | Schedule | Selected Prior Projects

Syllabus (printable) | Who / When / Where | Description | Prereqs | Grading | Topics

Who / When / Where

Instructor: Bradley Malin
Semester: Spring 2016
Time: Mondays & Wednesdays, 3:10 - 4:25pm
Location: Featheringill Hall, Room 313
Office Hours: *Upon request*, Location: 2525 West End Avenue, Suite 1030 (map)

Course Syllabus (PDF), Evacuation Plan (PDF)

First Day of Class: January 11, 2016

Description

The integration of information technology into biomedical environments has enabled unprecedented advances in the collection, storage, analysis, and rapid dissemination of patient-specific data. Many organizations need to share data for various purposes, such as quality assurance, public health, and basic research. In today's complex networked environments, it is increasingly difficult to share biomedical data due to concerns about patient privacy and anonymity. The goal of this course is to introduce students to the computational challenges, as well as formal solutions, for data privacy in healthcare and biomedical environments. Data privacy is an interdisciplinary problem, so this course will touch on issues in computer science, law and policy, and biomedicine.

Data Privacy Course Extensions: Personnel & Timeline

Vanderbilt

You Chen, Ph.D.
Biomedical Informatics

Daniel Fabbri, Ph.D.
Biomedical Informatics & Computer Science

Bradley Malin, Ph.D.
Biomedical Informatics, Biostatistics & Computer Science

UCSD

Xiaoqian Jiang, Ph.D.
Biomedical Informatics

Lucila Ohno-Machado, M.D., Ph.D.
Biomedical Informatics & Medicine

Shuang Wang, Ph.D.
Biomedical Informatics

T15 Supplement – Data Privacy (http://hiplab.mc.vanderbilt.edu/courses/BMIF380/)

Current Course

- Philosophy, Law, Policy, & Ethics
- Access Control and Auditing
- NLP for de-identification
- Re-identification methods
- Formal anonymization algorithms
- Advanced anonymization (e.g., differential privacy)
- Secure multiparty computation
- Case studies in biosurveillance, genomics, and medical record linkage

Extensions

- Electronic Consent
- Case studies based on the iDASH Privacy Challenge / Workshops (2015, 2016, & 2018)
- Secure hardware for biomedical data analytics
- Privacy preserving distributed data analytics (e.g., grid logistic regression)
- Migration to an online environment

T15 Supplement:
Case Studies as Educational Tools

- Introduce the complexity and multidisciplinary nature of biomedical informatics projects
- Present emerging issues and developing methodologies in evolving areas of the field
- Provide opportunities to identify social and ethical issues about health data and technology
- Illustrate generalizable lessons, which can be applied to other problems in the field

Case Studies as Educational Tools: Curriculum and Educational Goals

- Formalize the structure of biomedical data science case studies
- Create an initial series in a variety of open source formats, including print, online text, and video lectures
- Make case studies available for sharing across NLM training programs

Case Studies as Educational Tools: Proposed Cases

Case Study	Faculty
Building the infrastructure to support the precision medicine initiative	Paul Harris, Ph.D. Joshua Denny, M.D., M.S.
Learning new computational phenotypes from population-scale clinical data	Thomas Lasko, M.D., Ph.D.
Predicting suicide attempts in electronic health records	Colin Walsh, M.D., M.A.
Explanation-based electronic medical record auditing	Daniel Fabbri, Ph.D.
Learning organizational structure and collaborative teams through electronic medical record utilization	You Chen, Ph.D. Nancy Lorenzi, Ph.D.
Big data and why privacy does not have to die	Brad Malin, Ph.D.

Case Studies as Educational Tools: Personnel and Timeline

- Case Series Organizers and Course Directors:

You Chen, Ph.D.
Biomedical Informatics

Gretchen Purcell Jackson, M.D., Ph.D.
Biomedical Informatics & Surgery

Colin Walsh, M.D., M.A.
Biomedical Informatics & Medcine

- Project Manager: Carolyn Diehl
- Timeline:
  - Case Structure – February 2018
  - First Cases – April 2018
  - Course Offering – Summer 2018

Supplemental Training (Selected)

- Biomedical data science meet-ups once a month
  - Presentations by faculty and students on finished and in-progress research
  - Less formal setting to encourage interactive learning

- Pizza and High Performance Computing
  - http://www.accre.vanderbilt.edu/?page_id=3243
  - Monthly seminar on a range of topics
    - Python and analytics on HPC
    - Parallel processing in R
    - Big data tools and Hadoop
    - Running next gen sequencing on HPC

Data Science at Vanderbilt

- No official program

- Provost funded a transinstitutional program (TIP) to build a data science community
  - 5 Post-docs sponsored at 0.2 FTE
  - Standalone and joint-sponsored seminars
  - Think tanks (where specific datasets are reviewed)
  - University-wide conference

- Provost has convened two committees to develop strategic plan for
  - University-wide initiative (*includes Malin*)
  - M.S. program in data science, with application focus areas such as biomedicine (*includes Blume and Malin*)

Questions?

b.malin@vanderbilt.edu

Health Data Science Center
https://medschool.vanderbilt.edu/heads/

Health Information Privacy Laboratory
http://www.hiplab.org/

Center for Genetic Privacy & Identity in Community Settings
https://medschool.vanderbilt.edu/getprecise/