



UNC
ESHELMAN
SCHOOL OF PHARMACY

Data Science at the UNC Eshelman School of Pharmacy

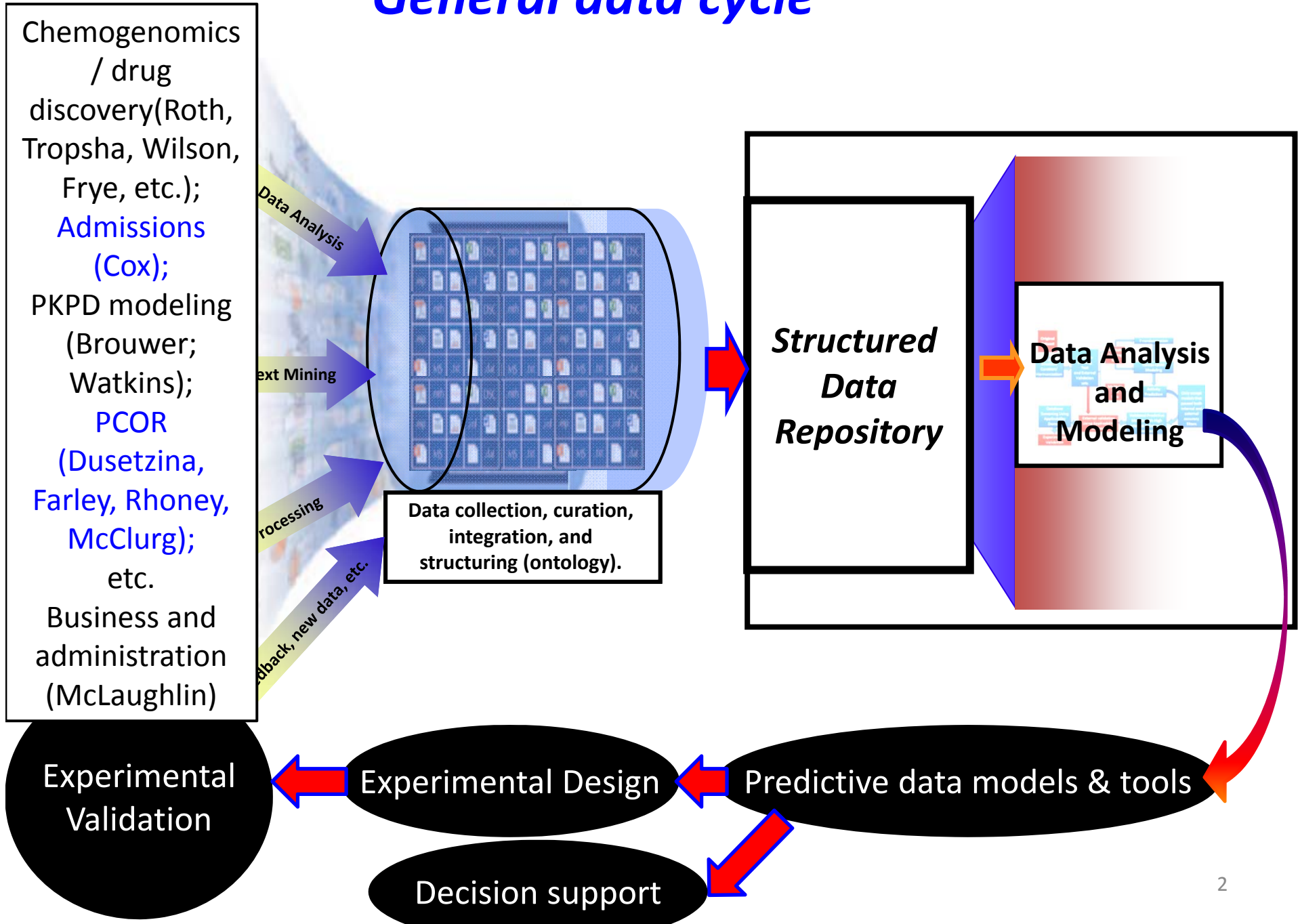
Alexander Tropsha

Associate Dean for Pharmacoinformatics and Data Science

UNC Eshelman School of Pharmacy

University of North Carolina, Chapel Hill, USA

General data cycle



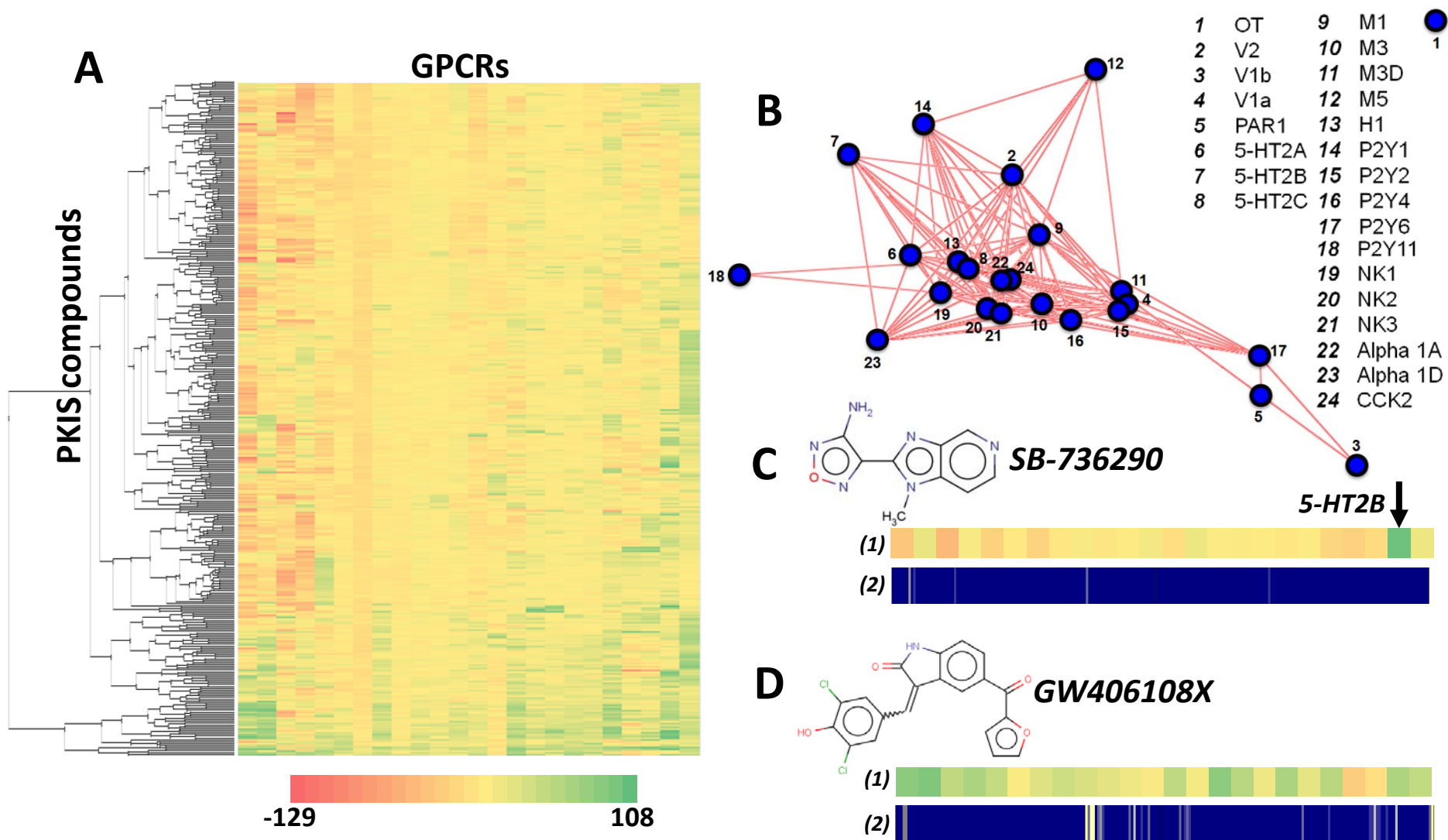


Domains of knowledge: Drug discovery, development, and care delivery.

| Process | Drug Discovery Development & Evaluation Care Delivery & Practice | | | | |
|---------------------------|---|--|--|----------------------------------|--|
| Academic Divisions | Medicinal Chemistry & Natural Products | Molecular Pharmaceutics | Pharmacotherapy & Experimental Therapeutics | Pharmaceutical Outcomes & Policy | Pharmacy Practice & Experiential Education |
| Centers | Center for Integrative Chemical Biology & Drug Discovery | Center for Nanotechnology in Drug Delivery | Center for Pharmacogenomics & Individualized Therapy | | |
| | | | Center for Medication Optimization (CMOP) | | |
| | | | Center for Simulations in Pharmacy (PharmSim) | | |

Data domains

- **Chemical genomics/drug discovery (CICBDD)**
 - 200,000 compounds stored in an Oracle/Biovia Direct database
 - 3,700,000 single dose screening results; 87,000 XC50's
- **Nanodelivery (CNDD)**
 - Nanomaterial Registry (ca. 2000 MNPs)
- **Pharmacogenomics (CPIT)**
 - The collection, curation, and storage of “big data” (e.g. genomic sequences for a patient cohort require terabytes of data) to investigate the interplay between genetic variation and drug efficacy/toxicity phenotypes
- **Medication compliance/optimization (CMOP)**
 - Multi-state Medicaid data ON CA. 100,000 patients with multiple chronic conditions
 - NC Medicaid data: 1.4M NC patients
 - Problems solved: the effect of pharmaceutical policies and interventions on medication access; Compare the effectiveness of different medications; examine practices that best improve patient care
- **Clinical Pharmacy**
 - Clinical data related to medication optimization in the intensive care unit; pharmacokinetic data.
 - Lack of big datasets needed to answer the complex questions in critical care.
- **Administration (Office of Strategic Planning and Assessment):**
 - Data collection to guide benchmarking, data reporting, and decision-making;
 - Data collection and annual reporting of faculty activity and productivity;
 - The administration, coordination, monitoring, and use of surveys and survey data conducted within the School;



PKIS – GPCR antagonist assay screening. (A) Heat map of responses with 24 clustered columns-GPCRs (X-axis) and 367 clustered PKIS compounds (Y-axis). (B) ADDAGRA representation of GPCRs. (C) and (D) GPCR antagonist profile (1) and kinase inhibition profile (2) are shown for selective SB-736290 and promiscuous GW406108X compounds.

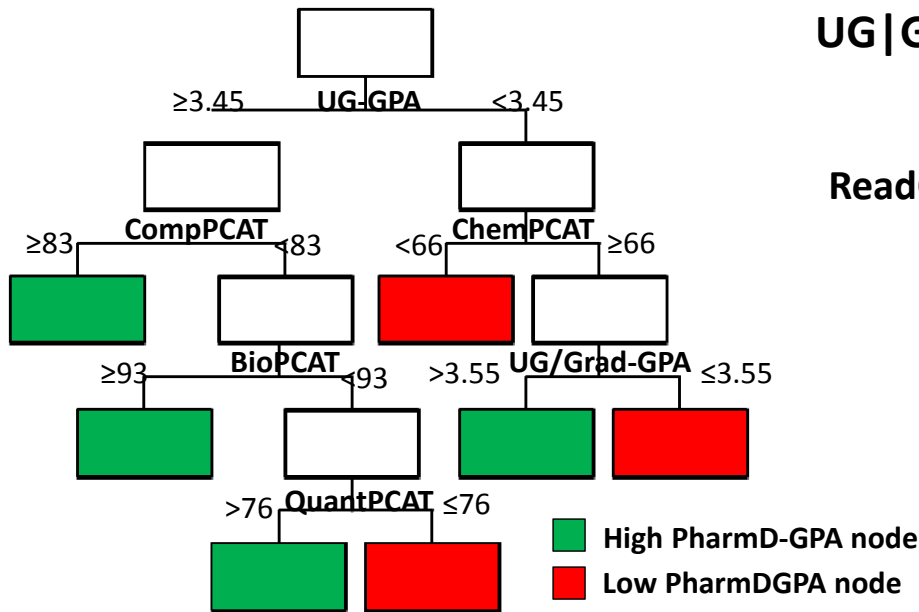
In collaboration with Drs. W. Zuercher and B. Roth

Statistical Modeling of Admission Data to Predict Academic Performance

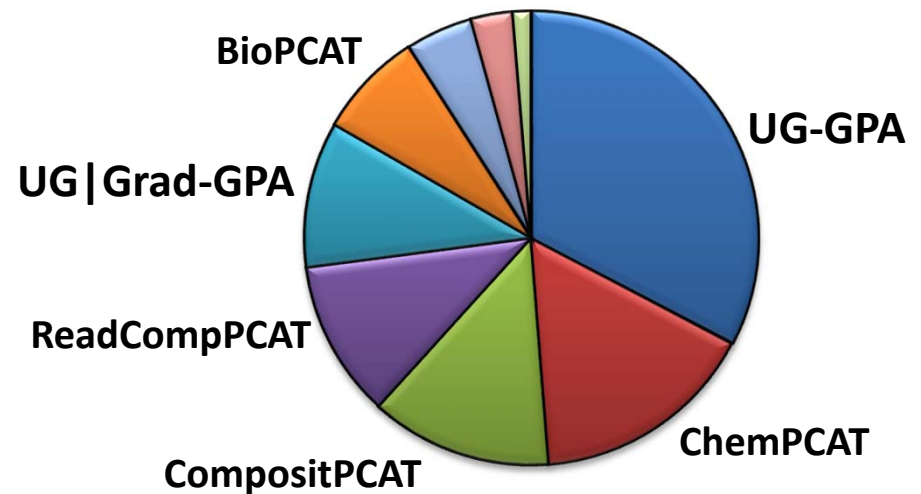
External Prediction Performances of RF models

| | BEST students (GPA>3.8) | WORST students (GPA<3) |
|---------------------|----------------------------|---------------------------|
| Prediction Accuracy | 92% | 95% |
| Coverage | 63% | 54% |

Example of decision tree



Contributions of parameters



In collaboration with Dr. W. Cox,
Assistant Dean for Professional Education

Vision for Data Science

- High-level Vision.

- Create sustainable DS infrastructure (hardware, software, domain-specific solutions) as part of the worldwide data science ecosystem to achieve data science driven solutions to societal problems

- education from K-12 through graduate and continuing education),
- health care,
- public health
- transportation,
- public safety,
- scientific discovery).
- Goal: Identify challenges where UNC-CH could make unique and impactful contributions by developing data-driven decision support tools and best practices.

- Campus-level vision

- Establish campus-wide, overarching Program in Data Science that develops and implements hardware (storage, wearable devices, computing platforms e.g. GPU, clusters etc.), software and methods (special algorithms for big data analytics), and methodological solutions (e.g., text mining/cognitive computing) to challenges faced by all activity domains
- Develop novel, technology driven educational tools
- Develop optimized health care delivery (novel tools to exploit Carolina Warehouse and similar EHR databases) including medication optimization and health outcomes driven clinical practices