

Data Sciences Hub

Brian Yandell

incoming Director, Biometry Program

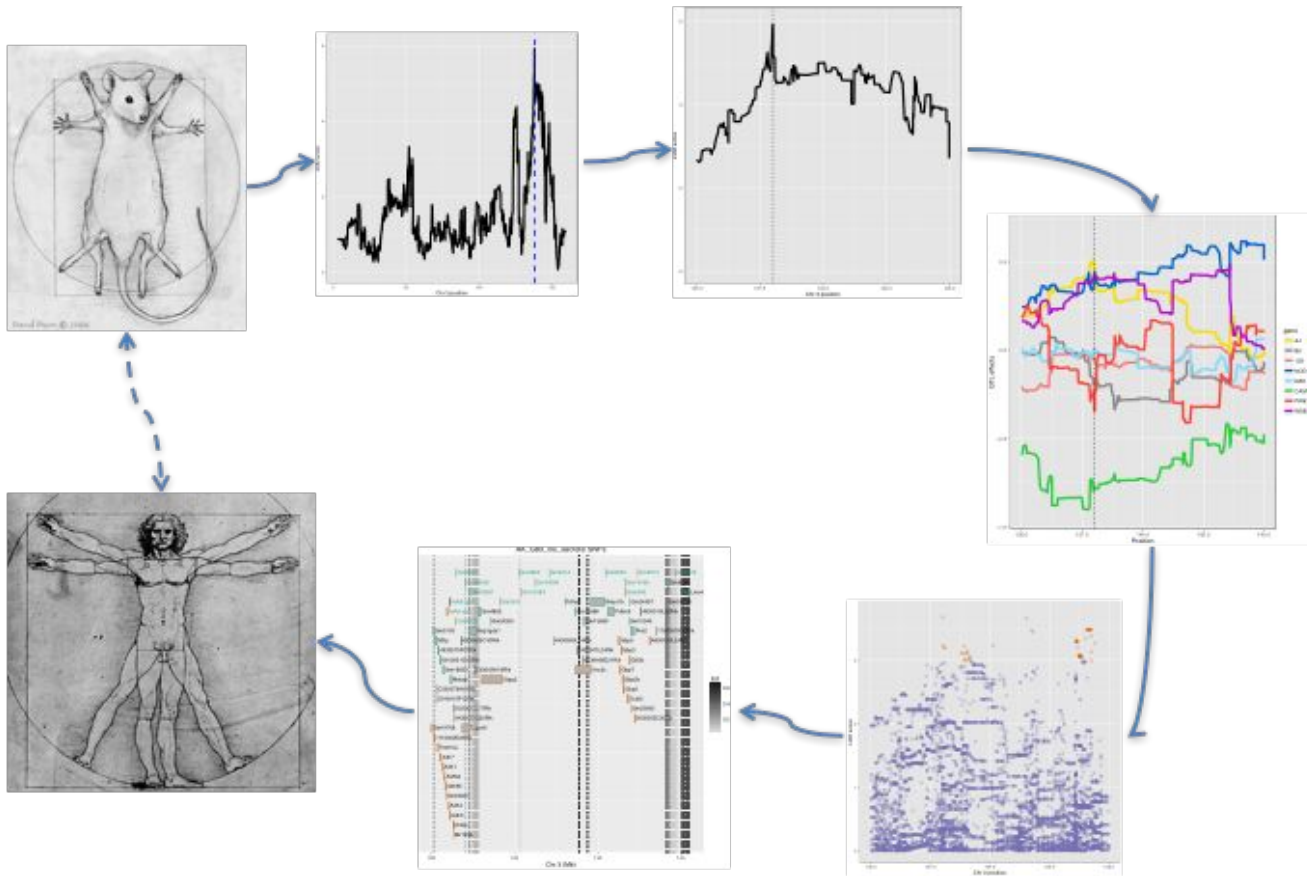
Biometry is "the active pursuit of ... knowledge by quantitative methods ... [with] constant experience in analysing and interpreting biological data.... [W]e come to think of ourselves ... in terms of the community of our interests with those doing similar work in other departments."

Sir RA Fisher (1948) inauguration of The Biometric Society

where is the campus not meeting data sciences needs?

- **Campus-wide consulting access** to plan & design data projects
 - statistical design to conceive & plan study
 - computational building blocks
 - engineering scale-up to deploy products
 - artistic visualization and communication of findings
- **Community of practice** across data sciences
 - inclusive access for all students, faculty & staff
 - training in data literacy & data competencies
 - shared teaching modules across domains
 - diversity of thought to tackle today's wicked problems
- Translating ideas into **reusable** pipelines & **deployed** apps
 - automate key steps to catalyze research collaborations
 - produce apps to share results broadly

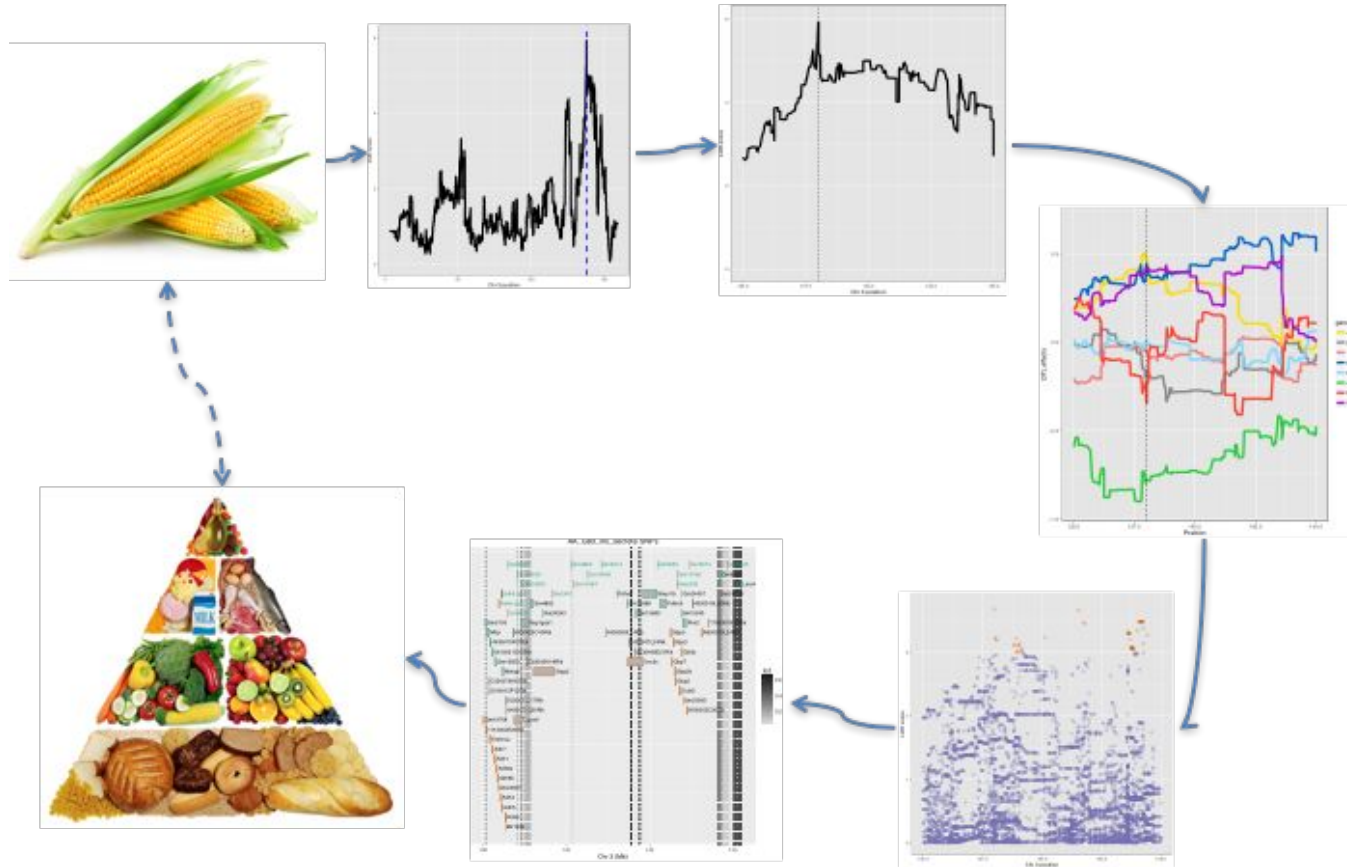
mouse genome team research pipeline



- Attie lab (Biochem)
- Broman lab (BMI)
- Kendziorski lab (BMI)
- Keles lab (BMI)
- Rey lab (Microb Sci)
- Coon lab (Chem)
- Yandell lab (Biometry)

- Churchill lab (Jax, ME)
- phenome.jax.org
- Williams lab (UT Memphis)
- Prins lab (Groningen, Neth)
- genenetwork.org
- Oak Ridge NL

maize genome team research pipeline



Kaepler lab (Agronomy)
Spalding lab (Botany)
de Leon lab (Agronomy)
WI Crop Innovation Ctr

Broman lab (BMI)
Yandell lab (Biometry)

what is needed to move data sciences to the next level?

- Create **campus-wide access** to collaboration
 - recognize added value of existing consulting & collaboration facilities on campus
 - leverage these to provide broader access
 - develop campus infrastructure & community of practice
- Foster **culture of research integrity** in team science
 - emphasize (early) planning and design
 - understand data in context of each project
 - be careful about misleading shortcuts
 - not simply access to new analytics toolbox
 - more data does not always mean better science

how could a data sciences hub help that move?

- Build competencies through training process
 - data & software carpentry
 - workflow & pipeline connectivity
 - communication in teams
- Address needs of various audiences
 - DSHub integrators/automators/collaborators
 - campus staff & graduate students in team research projects
 - teams from industry or government
- Who will do training?
 - leverage existing courses & modules (from UW or online)
 - help build customized material in small snippets/modules
- Basics are free; added value generates revenue