

---

# The CompGen Platform

Ravi Iyer & Steve Lumetta

Dept. of ECE, Coordinated Science Lab  
University of Illinois at Urbana-Champaign

---

# Today's Talk

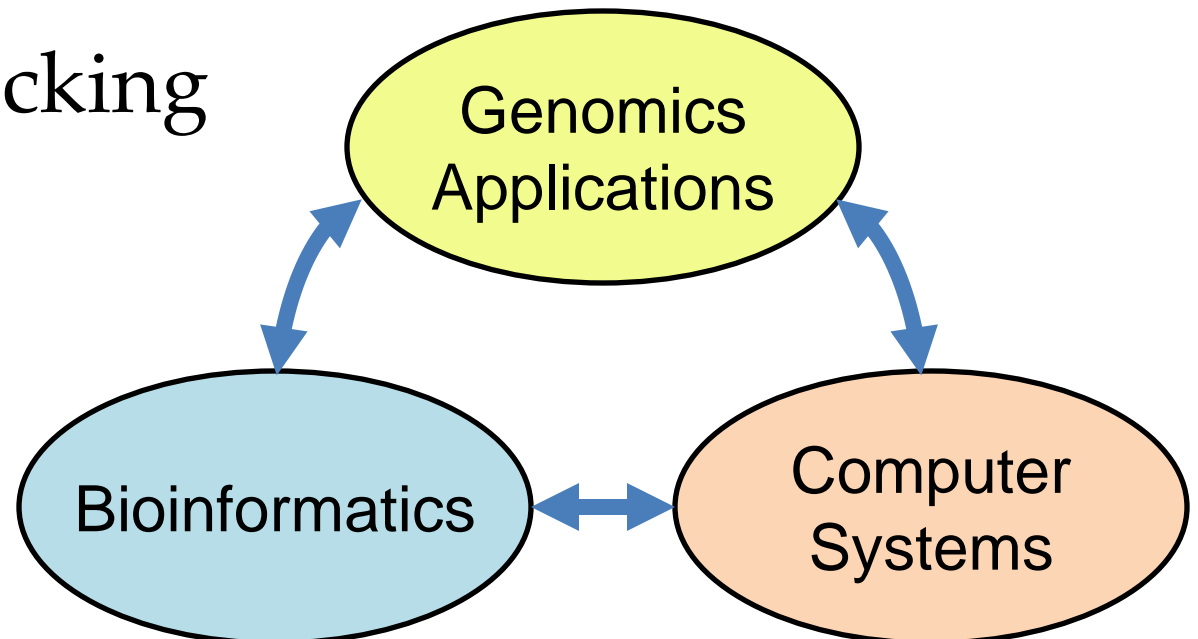
---

- overview (Ravi)
- vision: CompGen instrument as a catalyst
- Why now?
- some initial research targets
- computer systems opportunities
- organization plan

# Catalyze Computational Genomics

---

- create experimental platform housed in IGB
- leverage state-of-the-art technology
  - high-speed network, GPUs, FPGAs
  - non-volatile RAM
  - 3D die-stacking (memory)



# Interdisciplinary Skills are Valuable

---

- Mendel (1865-70):  
combinational math predicts existence of genes
- Avery (1944):  
bacterial transformations of medical interest lead to identification of DNA as carrier of genes
- Watson & Crick (1953):  
Franklin & Gosling's X-ray crystallography uncovers helical DNA structure

# Domain-Specific Computing is Important

---

Successful targeted areas have included...

- logarithms and arithmetic
- codebreaking
- systems of linear equations
- artillery trajectories/firing tables
- bubble chamber image processing
- graphics processing (GPUs)
- protein folding
- molecular dynamics

# Benefits of Collaboration Demonstrated

---

- CSL founded in 1951 for interdisciplinary research (then on control systems)
- IGB opened in 2007 to unite biotech research at Illinois
- ca. 2011, Gene & Ravi formed a cross-campus group of biologists and computing researchers
- examples of team benefits already shown
  - Chen, Hwu, Ma: de novo assembly
  - Campbell, Farivar, Jongeneel: GPU-based sequencing
  - Iyer, Robinson: phylogeny reconstruction

# Why Am I Excited?

---

(computing perspective)

- need
  - domain provides metrics for success
  - real-world applications
  - inspire young minds with challenging and important problems
- opportunity
  - data volume enables new models, analyses
  - models and analyses offer opportunities
  - expect lots of changes, and many advances

# Why Should a Biologist be Excited?

---

- need
  - vast amounts of heterogeneous data (laptop model failing)
  - many now run on clouds
    - good for some applications
    - can also tune for some applications (e.g., Convoy)
  - but does cloud does not support well
    - changes to commercial model need clear motivation
    - data will evolve rapidly in near future
- opportunity
  - processor power wall ~2004
  - limited gains in single processor performance
  - increasing number of special-purpose accelerators on chip / package
  - Why not some for genomics?

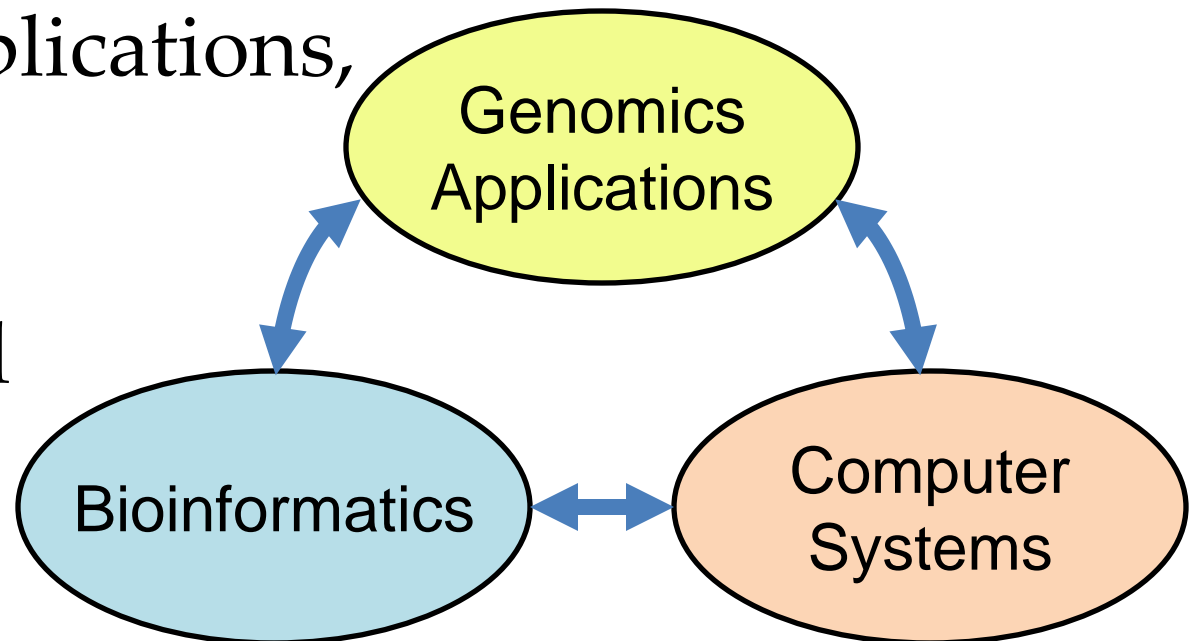


# MRI: We'll Develop an Instrument

---

- (US) National Science Foundation
- Major Research Instrumentation program
- 4-year project to develop instrument
- enable research collaborations

spanning applications,  
work flows,  
algorithms,  
statistics, and  
technologies

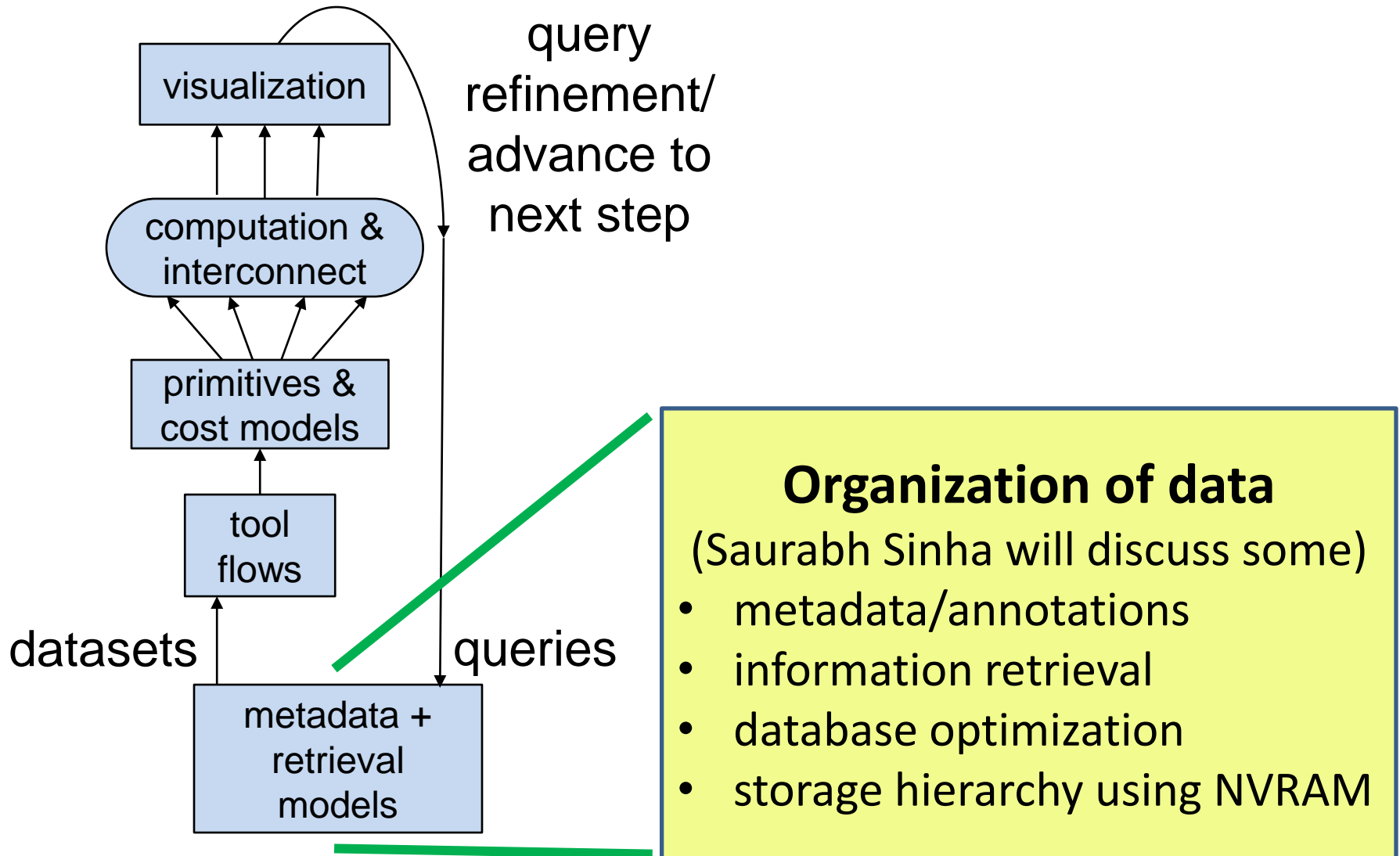


# Some Research We Want to Enable

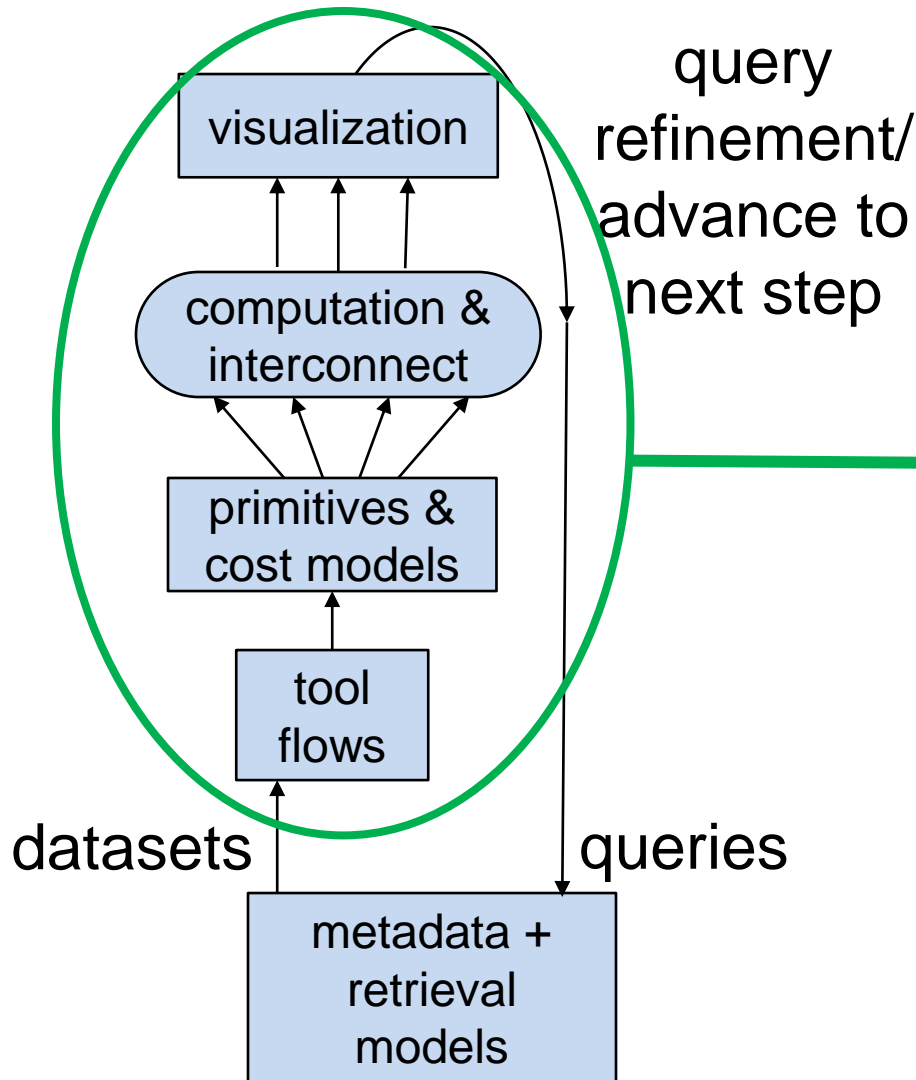
---

- accurate detection of genomic variants
- metagenomics workflows, algorithms, and statistics
- acceleration of phylogeny reconstruction
- drug discovery from microbial genomes

# Computing Challenges Span R&D



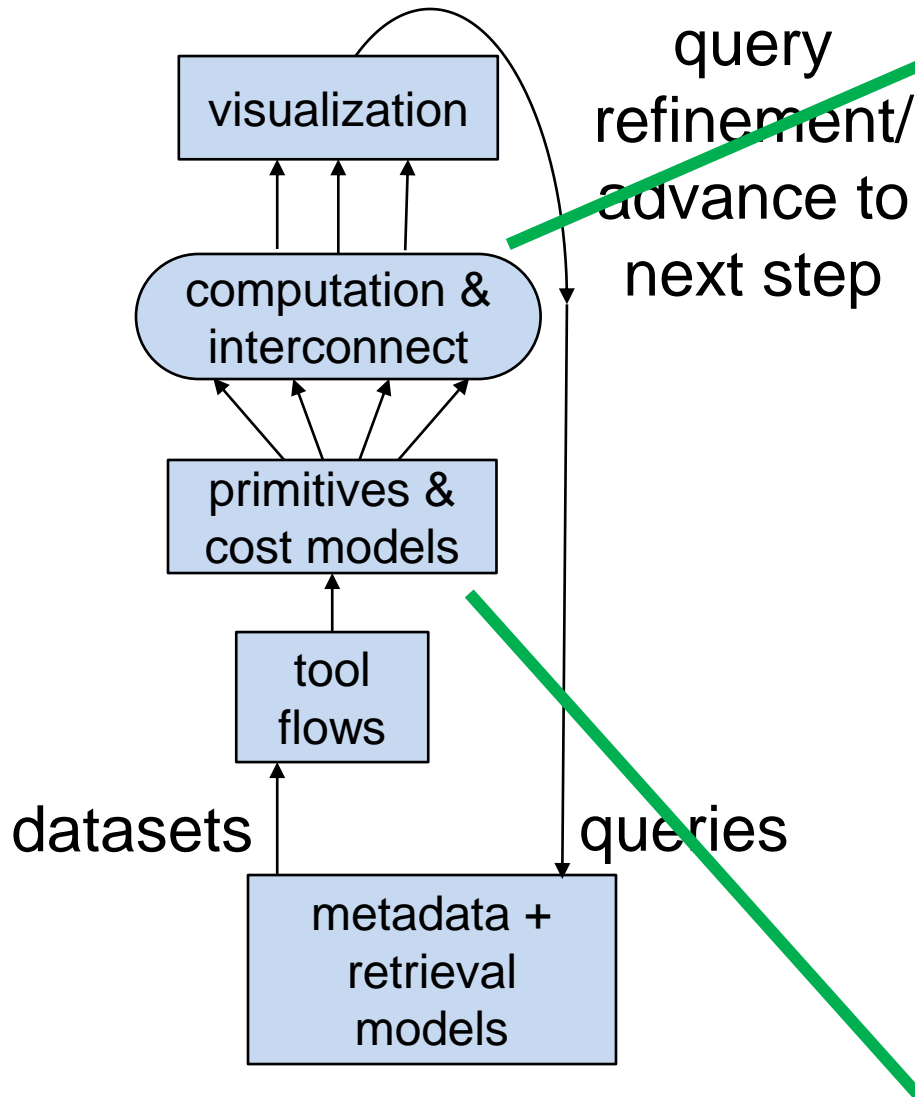
# Optimize Data Motion



## Management of Dataflow

- coordinated resource scheduling
- reduction of overhead
- data reduction, execution on compressed data
- stacked DRAM, memory caching
- visualization

# Identify New Computational Strategies



## Execution of Computation

(Victor Jongeneel & Wen-mei Hwu will discuss some)

- identification of primitives
- emerging/evolving GPU architectures
- acceleration and prototyping with FPGAs
- programming environment
- cost models and dynamic algorithm selection

# Mgmt. Oversees Development and Use

---

- Project Management Team  
(Lumetta, Iyer, Jongeneel, Robinson, Sinha)
  - oversee development process & team
  - coordinate consortium meetings
  - manage risks
- Research Steering Committee  
(PM + Rodriguez-Zas, Stubbs, Winslett)
  - encourage collaborative use proposals
  - oversee allocation of instrument
- +27 other faculty members

# 3-Person Team Drives Development

---

- **Instrument Utilization Manager**  
day-to-day operations and oversight
- **Software Engineer**  
develop toolkits, system monitoring, tools,  
and optimization strategies
- **Hardware Engineer**  
develop and integrate h/w technologies

# Transition to Research Over Time

---

- target post-doctoral candidates
- initial buildout (first two years)
  - focus on identification and development of technologies
  - work with research project teams to enhance instrument use
- last two years
  - development effort scales down
  - transition part time to research projects



# Bring the Community Together

---

- Consortium of 15 institutions
  - formed in conjunction with MRI effort
  - committed to coming here twice / year
  - distributed across areas
- Bioinformatics: Agilent, Strand
- Computing: IBM, Infosys, Intel, Microsoft, MulticoreWare, NVIDIA, TIFR, Tezzaron
- Genomics: Abbott, Baylor, BGI, Mayo, Monsanto, Wash. U.

# Good News

---

- MRI will be funded!
- Starting in Fall 2013
- Looking for good development team candidates!

# Plan to Meet Late Fall 2013

---

(kick-off meeting for Consortium)

- a few high-level talks by faculty
- grad students present to members
  - progress reports on research
  - thinking on next directions
- plenty of time for socialization
  - get feedback from members
  - opportunities to form new collaborations