

# NHGRI Computational Genomics and Data Science Program Workshop: Overview and Recommendations

Ajay Pillai  
on behalf of CGDS Group

NHGRI Council  
May 8, 2017

# NHGRI CGDS Workshop: Rockville, September 29-30 2016

---

## Goals

- Prioritize genomic research topics relevant to the NHGRI extramural Computational Genomics and Data Science (CGDS) program
- Identify current challenges that face the computational genomics and data science community
- Redefine the focus of the CGDS portfolio over the next 3-5 years

# NHGRI CGDS Workshop

---

## Meeting Organizers

- Drs. Mike Boehnke, Carol Bult, Trey Ideker, Aviv Regev, Lincoln Stein
- NHGRI CGDS Staff

## Participants

- 39 invited extramural researchers (academic and industry)
- NHGRI, NCI, NIGMS, ADDS staff

# NHGRI CGDS Workshop

---

## Organization

- Sessions designed by organizing committee
- Speakers selected by small groups of participants

## Session Topics

- Challenges in enabling new biology in basic science
- Challenges in enabling new clinical insights
- Data and computational resources
- Computations at scale
- Collaborating with non-NHGRI resources

# NHGRI CGDS Workshop

Recommendations  
from sessions  
consolidated and  
refined by  
participants using  
dot-storming

The screenshot shows a dot-storming interface for a workshop. At the top, it says "dotstorming" and "NHGRI Informatics/Data Science Priorities". Below this, instructions state: "You have 6 'dots' to use for voting. You can use one dot per topic or use multiple dots for one topic. Once you vote 6 times you'll be out of dots!" and "\*\*You can also add comments to the topics." On the right, there is a box to "Enter your name to participate" with a "Join" button. The main area contains several topic cards, each with a title, a description, and a row of 6 dots for voting. Some cards have comments. A "Participants 32" list is on the right side.

**23. Invest heavily in visualization tools, as a well defined goal**  
I'VE BEEN INCORPORATED INTO #23. DON'T WASTE A DOT ON ME.

**5. NHGRI support rigorous benchmarking and, where possible/necessary, development of "gold standards"**

**19. Support development of statistical and computational methods and tools to identify causal variants, likely taking advantage of existing available data, such as well phenotyped and genotyped individuals and functional annotations.**

**17. Support development of statistical and computational tools that enable interactive analysis of large datasets to enable thoughtful balancing of human expert input with machine automation. Avoid building pipelines/workflows prematurely.**  
23. Invest heavily in visualization tools, as a well defined goal

**18. Support development of statistical models and computational methods that account for stochastic behavior in biology, for example, for single-cell data.**  
I'VE BEEN COMBINED WITH #3. DON'T WASTE A DOT ON ME!

**15. NHGRI should promote the FAIR use of software tools**

**10. Promote the development and use of data and service interoperability standards.**

**7. NHGRI should invest in resources focused on curated metadata and computable phenotypes**  
THE METADATA COMPONENT IS NOW IN #20. THE PHENOTYPES COMPONENT IS #25.  
DON'T WASTE A DOT ON ME.

**Participants 32**

Adam Arkin
Angel Pizarro
Angela Brooks
Anne Kwitek
Anthony Philippakis
Aviv Regev
Bob Grossman
Carol Bult
Casey Overby
chiara sabati
Corrie Painter
David Glazer
Ekta Khurana
Hector Corrada Bravo
Hyun Min Kang
James Taylor
Jessie Tenenbaum
Jyoti Pathak
Lincoln Stein
Melissa Gymrek

# 13 Recommendations for NHGRI extramural support of CGDS

---

1. Statistical/computational tools enabling ***interactive analysis and visualization of large data sets***
2. Methods and data that enhance ***understanding of how genotype translates to phenotype***
3. Tools, technologies, and policies to ***ensure genomic data sharing***
4. Statistical/computational tools to ***identify causal variants***

# 13 Recommendations for NHGRI extramural support of CGDS

---

5. Multi-scale *phenotype-focused ontologies* and standards
6. *Efficient and scalable algorithms* and methods for compute-intensive applications
7. *Vertically integrated data resources* supporting horizontally-organized knowledgebases
8. Methods enabling *scalable, intelligent, cost-effective FAIR\* metadata*

FAIR\*: Findable, Accessible, Interoperable, Reusability

# 13 Recommendations for NHGRI extramural support of CGDS

---

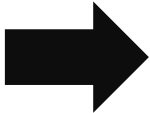
9. ***Cloud environment*** for NHGRI investigators to share data and tools
10. ***Rigorous benchmarking*** and development of ‘gold standards’
11. Improvements needed to ***integrate genomic medicine into Clinical Decision Systems***
12. ***Integrating patients more fully into genomic medicine research*** and clinical practice
13. Support informatics and computational needs for ***single cell studies***



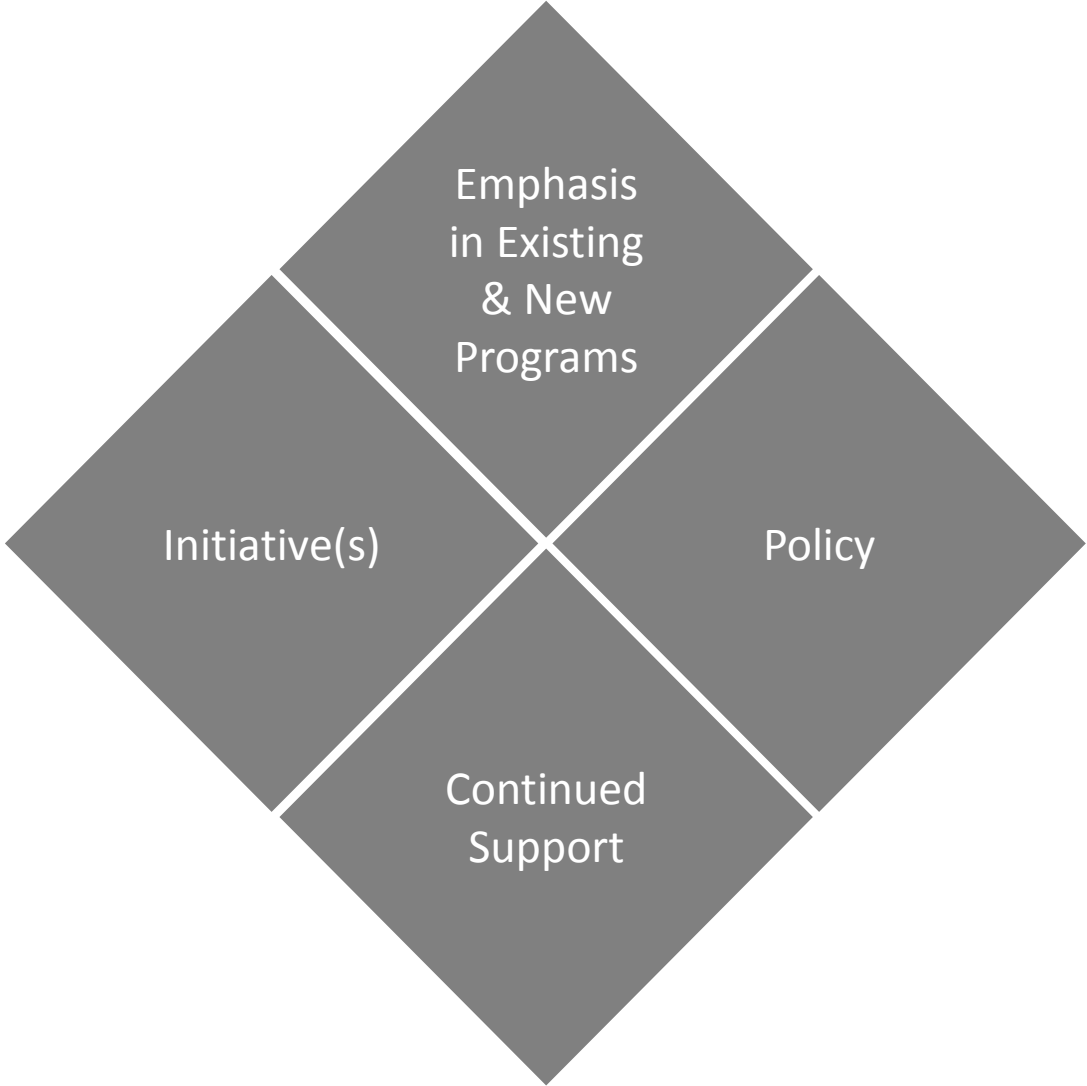
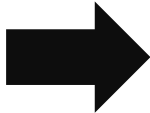
# Moving from Recommendations to Potential Actions

---

**Workshop  
Recommendations**



**Portfolio  
Analysis**

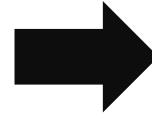


# Example Portfolio Analysis: Visualization (Interactive)

---

**1 - Create query and aliases  
(avoid false negatives)**

(e.g., “genome visualization”~10 OR  
“variant browser”~10 OR “population  
variation visualization”~10)



**2 – Place constraints**

(e.g., number of years, awards, non-awarded, HG-only, non-HG)



**3 – Curate results  
(remove false positives)**

Remove hits



# Example Portfolio Analysis Results: Visualization (Interactive)

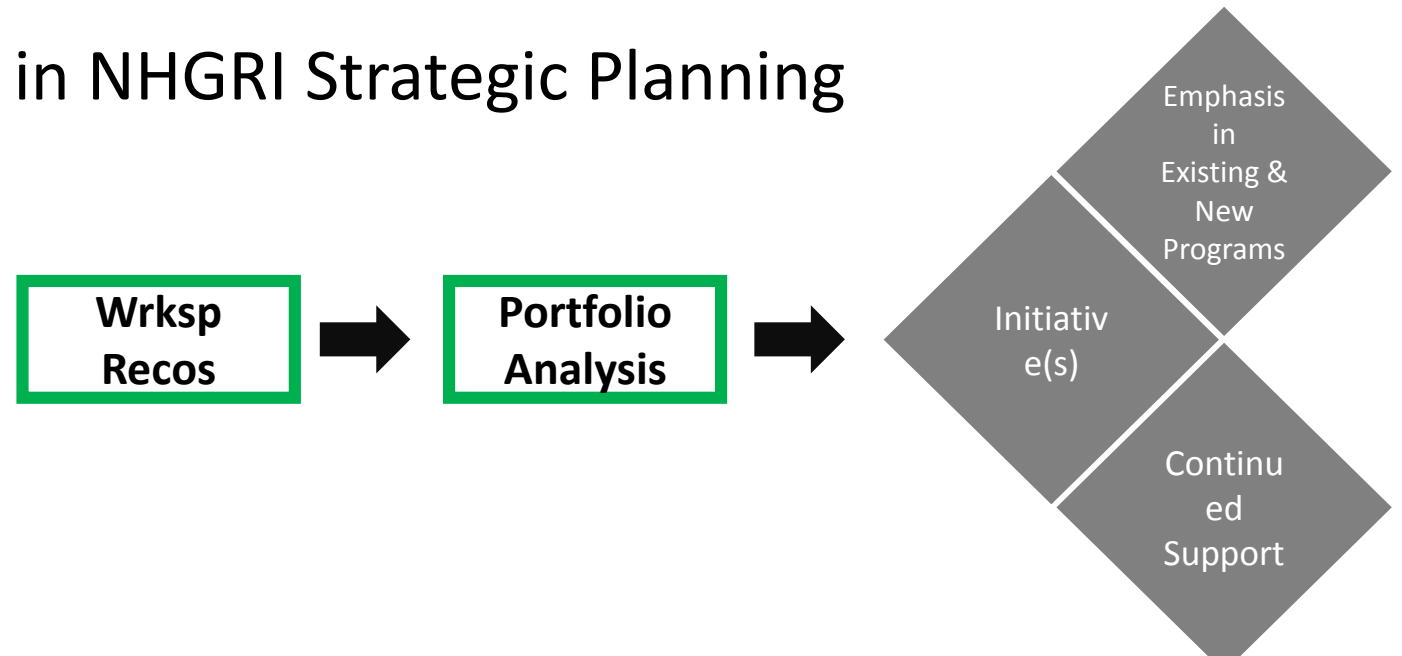
---

Query	HG funded (before; after curation) last decade	Non-HG funded (before; after curation) last decade
“genome visualization”~10	42;11	131;42
“genome visualization”~10 AND interactive	3;3	27;14

- Query runs on abstract and specific aims.
- We can obtain detailed information on both awarded and non-awarded grants and potentially utilize the results to correctly reflect the needs.

# Next Steps

- Publish report on website and advertise
- Expect to finish portfolio analysis by summer
- These will be used as input in NHGRI Strategic Planning



# Acknowledgements and Questions

---

The co-chairs:

Drs. Mike Boehnke, Carol Bult, Trey Ideker, Aviv Regev, Lincoln Stein

Kevin Lee

NHGRI CGDS Staff: Lisa Brooks, Valentina Di Francesco, Dan Gilchrist, Mike Pazin, Erin Ramos, Heidi Sofia, Jen Troyer, Chris Wellington, Ken Wiley.

Eric Green, Carolyn Hutter, and Jeff Schloss,