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

Ajay Pillai
on behalf of CGDS Group

NHGRI Council
May 8, 2017
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
Recommendations from sessions consolidated and refined by participants using dot-storming.
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*
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
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

- Emphasis in Existing & New Programs
- Initiative(s)
- Continued Support
- Policy
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)

<table>
<thead>
<tr>
<th>Query</th>
<th>HG funded (before; after curation) last decade</th>
<th>Non-HG funded (before; after curation) last decade</th>
</tr>
</thead>
<tbody>
<tr>
<td>“genome visualization”~10</td>
<td>42;11</td>
<td>131;42</td>
</tr>
<tr>
<td>“genome visualization”~10 AND interactive</td>
<td>3;3</td>
<td>27;14</td>
</tr>
</tbody>
</table>

- 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
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,