

**U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES  
NATIONAL INSTITUTES OF HEALTH (NIH)  
NATIONAL LIBRARY OF MEDICINE (NLM)  
BOARD OF REGENTS (BOR) COMPARATIVE GENOMICS RESOURCE (CGR)  
WORKING GROUP (WG) MEETING  
January 23, 2024**

**MEMBERS PRESENT**

Kristi L. Holmes, PhD, Northwestern University, Chair

**EXTERNAL MEMBERS PRESENT**

Ani W. Manichaikul, PhD, University of Virginia School of Medicine  
Terence Murphy, PhD, NCBI, Executive Secretary  
Len Pennacchio, PhD, Lawrence Berkeley National Laboratory  
Kenneth Stuart, PhD, Seattle Children's Research Institute  
Tandy Warnow, PhD, University of Illinois, Champaign-Urbana  
Rick Woychik, PhD, NIEHS & NTP, NIH CGR Steering Committee Liaison  
Cathy Wu, PhD, University of Delaware

**EXTERNAL MEMBERS NOT PRESENT**

Alejandro Sanchez Alvarado, PhD, Stowers Institute for Medical Research  
Hannah Carey, PhD, University of Wisconsin-Madison

**OTHERS PRESENT**

Anne Ketter, Sr. Product Manager, NCBI  
Nuala O'Leary, PhD, Product Owner, NCBI Datasets, NCBI  
Valerie Schneider, PhD, NCBI  
Maureen Sides, Staff Assistant, NIEHS  
Ekaterina Sukharnikov, PMP, Project Manager, NCBI

## **I. WELCOME AND INTRODUCTIONS**

*Kristi Holmes, PhD, Chair*

Dr. Holmes greeted Working Group members and thanked them for their continued participation in the Working Group. She introduced the agenda topics for the meeting and welcomed the Working Group members' continuing feedback on how best to engage the scientific community to promote CGR.

## **II. CGR PROJECT UPDATES**

*Terence Murphy, PhD, NCBI, Executive Secretary*

*Kristi Holmes, PhD, Chair*

*Working Group Members*

Dr. Murphy introduced himself as the new Project Lead for the CGR project and provided an overview of recent progress. The Foreign Contamination Screen (FCS) tool, FCS-GX, is now available via the web-based Galaxy analysis platform, offering increased access to the tool. A revised manuscript detailing the FCS tool was also submitted. Development of the EGAPx genome annotation pipeline is also in progress, with early alpha testing anticipated to begin in the near future. NCBI has also continued work to provide additional data analysis and connections for users, including additional gene functional information from more than 1750 eukaryote species provided via a Gene Ontology data section. In addition, NCBI has rebuilt a matching pipeline to help users compare information between the NCBI Reference Sequence (RefSeq) and Universal Protein Resource (UniProt) databases. Dr. Murphy also noted additional progress in the Comparative Genome Viewer (CGV) and the BLAST ClusteredNR database. The CGR Public Roadmap was presented, including updates since August 2023 such as enhancements to the NCBI Datasets tool as well as other database improvements.

Dr. Murphy noted several channels through which CGR users can receive updates regarding the project, including contacting the CGR team at NCBI by email as well as subscribing to the CGR project newsletter, joining the mailing list, or visiting the website. WG members were also reminded to monitor for new and updated materials provided via the CGR BOR WG Teams site for use in engagement with research communities. Focused engagement activities, particularly targeting librarians, professors, and students, since September 2023, were highlighted. The CGR review white paper discussing the impact of comparative genomics on human health was published in *BMC Genomics* in September 2023, and additional manuscripts regarding CGR component tools and databases have also been submitted for publication.

Regarding the WG's engagement goals, Dr. Murphy noted the challenge of centrally tracking the WG members' outreach and engagement activities. He noted that NCBI-developed flyer or handout documents detailing the CGR project will be uploaded to the CGR BOR WG Teams site and distributed to the WG for sharing with their communities. To further

support engagement, any additional informational materials detailing recent updates or other information can be developed upon request. WG members evaluated, via polling, their knowledge of the CGR team's activities and outreach and their engagement with the CGR newsletter.

### **III. NCBI DATASETS DEMO**

*Nuala O'Leary, PhD, Product Owner, NCBI Datasets, NCBI Working Group Members*

Dr. O'Leary presented the NCBI Datasets tool, emphasizing its development through community engagement. The tool aims to facilitate easy navigation and use of NCBI data, particularly given the rapid explosion of available genomic data. The NCBI Datasets tool offers both web-based and programmatic interfaces, providing users with versatile access to NCBI data. Requested data and metadata can then be delivered together in the form of a data package based on findability, accessibility, interoperability, and reusability (FAIR) data principles. The web interface, which includes organism-focused pages and direct links to other NCBI databases, was highlighted.

Dr. O'Leary provided a demonstration of the NCBI Datasets tool, including how to browse for and download both genome and gene ortholog data. The NCBI Datasets Documentation pages were also highlighted, including How-to Guides, Tutorials, and Reference pages with additional information on metadata schemas and the NCBI Datasets application programming interface (API). Dr. O'Leary also noted the user feedback and citations, being used to evaluate the success of the NCBI Datasets tool.

Per a WG member's question, Dr. O'Leary noted the possibility of expanding search parameters within the NCBI Datasets tool to include additional comparisons between different groups of organisms. Additionally, via polling, WG members confirmed their understanding of the NCBI Datasets tool and acknowledged its potential impact within their research communities.

### **IV. CGR WORKFLOW ACTIVITY**

*Terence Murphy, PhD, NCBI, Executive Secretary*  
*Rick Woychik, PhD, NIEHS & NTP, NIH CGR Steering Committee Liaison*  
*Working Group Members*

Dr. Murphy presented the CGR Workflow activity, allowing WG members to help develop "short workflow" use cases to help promote CGR tools in both marketing and educational materials. Generic sample scenarios were presented, illustrating how specific research questions can be addressed using CGR tools. Dr. Woychik also noted the importance of identifying key capabilities or use cases that can help set CGR apart from other genome analysis resources.

WG members recommended highlighting the scope of genome data accessible via CGR to differentiate it from other resources which may be limited to data from more common organisms; members noted that many existing resources predominately focus on mammalian or vertebrate genome data. NCBI also confirmed that engagements with other existing genome data resources are ongoing to increase interconnectedness and data accessibility.

WG members also noted the greater depth and precision with which both comprehensive comparisons across diverse organisms and gene orthology and evolution information can be found using CGR tools, as opposed to other resources.

## **V. WG OUTREACH**

*Terence Murphy, PhD, NCBI, Executive Secretary*

*Rick Woychik, PhD, NIEHS & NTP, NIH CGR Steering Committee Liaison*

*Working Group Members*

Dr. Murphy facilitated a discussion of outreach efforts with the WG and highlighted the availability of the CGR BOR WG Communications Toolkit on the CGR BOR WG Teams site. Dr. Woychik recommended direct coordination between NCBI and the WG to ensure members are well-informed about prioritized outreach and engagement efforts. WG members also emphasized the need for increased development of educational materials to train future researchers in utilizing and understanding the benefits of CGR. WG members and NCBI will also continue to seek feedback from research communities to develop additional use cases and workflows for CGR tools.

Per WG members' recommendations, it was agreed that NCBI will explore implementing the distribution of periodic updates to the WG to outline recent CGR project developments which may be worth sharing with research communities. WG members also requested that NCBI recommend specific engagement activities which may be most coherent with the project updates. Dr. Woychik reiterated the importance of fostering greater adoption of CGR within the scientific community given its value for genomics research.

## **VI. WRAP UP**

*Kristi Holmes, PhD, Chair*

*Working Group Members*

Dr. Holmes thanked WG members for their continued participation and for their valuable feedback on outreach efforts and on use case development for CGR. NCBI will continue to provide WG members with resources and materials to help them promote CGR.