Accepted Manuscript

Title: Raising orphans from a metadata morass: a researcher's guide to re-use of public 'omics data

Author: Priyanka Bhandary Arun S. Seetharam Zebulun Arendsee Manhoi Hur Eve Syrkin Wurtele

PII: S0168-9452(17)30574-5
DOI: https://doi.org/doi:10.1016/j.plantsci.2017.10.014
Reference: PSL 9693

To appear in: *Plant Science*

Received date: 17-7-2017
Revised date: 7-10-2017
Accepted date: 15-10-2017

Please cite this article as: Priyanka Bhandary, Arun S. Seetharam, Zebulun Arendsee, Manhoi Hur, Eve Syrkin Wurtele, Raising orphans from a metadata morass: a researcher's guide to re-use of public 'omics data, <*![CDATA[Plant Science]]*> (2017), https://doi.org/10.1016/j.plantsci.2017.10.014

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Raising orphans from a metadata morass: a researcher's guide to re-use of public 'omics data

Priyanka Bhandary[a,b], Arun S. Seetharam[c], Zebulun Arendsee[a,b], Manhoi Hur[a,b], Eve Syrkin Wurtele[a,b,*]

[a] *Dept. of Genetics Development and Cell Biology, Iowa State University, Ames IA, 50010, USA*
[b] *Center for Metabolic Biology, Iowa State University, Ames, IA 50011, USA*
[c] *Genome Informatics Facility, Office of Biotechnology, Iowa State University, Ames, IA 50011, USA*

Abstract

More than 15 petabases of raw RNAseq data is now accessible through public repositories. Acquisition of other 'omics data types is expanding, though most lack a centralized archival repository. Data-reuse provides tremendous opportunity to extract new knowledge from existing experiments, and offers a unique opportunity for robust, multi-'omics analyses by merging metadata (information about experimental design, biological samples, protocols) and data from multiple experiments. We illustrate how predictive research can be accelerated by meta-analysis with a study of orphan (species-specific) genes. Computational predictions are critical to infer orphan function because their coding sequences provide very few clues. The metadata in public databases is often confusing; a test case with *Zea mays* mRNA seq data reveals a high proportion of missing, misleading or incomplete metadata. This metadata morass significantly diminishes the insight that can be extracted from these data. We provide tips for data submitters and users, including specific recommendations to improve metadata quality by more use of controlled vocabulary and by metadata reviews. Finally, we advocate for a unified, straightforward metadata submission and retrieval system.

*Corresponding Author.
Email address: mash@iastate.edu (Eve Syrkin Wurtele)

Accepted Manuscript

Download English Version:

https://daneshyari.com/en/article/8356782

Download Persian Version:

https://daneshyari.com/article/8356782

Daneshyari.com