OMICS A Journal of Integrative Biology
Volume 18, Number 6, 2014
© Mary Ann Liebert, Inc.
DOI: 10.1089/omi.2014.0061

Review Article

MOPED 2.5—An Integrated Multi-Omics Resource: Multi-Omics Profiling Expression Database Now Includes Transcriptomics Data

Elizabeth Montague,[1–4] Larissa Stanberry,[1–4] Roger Higdon,[1–4] Imre Janko,[2–4] Elaine Lee,[2–4]
Nathaniel Anderson,[1,2,4] John Choiniere,[1,2,4] Elizabeth Stewart,[1,4] Gregory Yandl,[1,3,4]
William Broomall,[2–4] Natali Kolker,[2–4] and Eugene Kolker[1–5]

Abstract

Multi-omics data-driven scientific discovery crucially rests on high-throughput technologies and data sharing. Currently, data are scattered across single omics repositories, stored in varying raw and processed formats, and are often accompanied by limited or no metadata. The Multi-Omics Profiling Expression Database (MOPED, http://moped.proteinspire.org) version 2.5 is a freely accessible multi-omics expression database. Continual improvement and expansion of MOPED is driven by feedback from the Life Sciences Community. In order to meet the emergent need for an integrated multi-omics data resource, MOPED 2.5 now includes gene relative expression data in addition to protein absolute and relative expression data from over 250 large-scale experiments. To facilitate accurate integration of experiments and increase reproducibility, MOPED provides extensive metadata through the Data-Enabled Life Sciences Alliance (DELSA Global, http://delsaglobal.org) metadata checklist. MOPED 2.5 has greatly increased the number of proteomics absolute and relative expression records to over 500,000, in addition to adding more than four million transcriptomics relative expression records. MOPED has an intuitive user interface with tabs for querying different types of omics expression data and new tools for data visualization. Summary information including expression data, pathway mappings, and direct connection between proteins and genes can be viewed on Protein and Gene Details pages. These connections in MOPED provide a context for multi-omics expression data exploration. Researchers are encouraged to submit omics data which will be consistently processed into expression summaries. MOPED as a multi-omics data resource is a pivotal public database, interdisciplinary knowledge resource, and platform for multi-omics understanding.

Introduction

Data-driven science is fueled by high-throughput technologies that produce large amounts of multi-omics data. The ability to access, compare, and analyze these data is critical for future scientific discoveries. Yet once the data are collected, they are often stored in raw or variously processed forms within scattered public repositories, accompanied by differing amounts of experiment metadata. Even more data remain in private storage, inaccessible and underutilized. Public repositories are generally separated by omics, which then complicates multi-omics analysis (Barrett et al., 2012; Desiere, 2006; Edgar et al., 2002; Maglott, 2004; Parkinson et al., 2011; Rebhan et al., 1997; Stelzer et al., 2011; Vizcaíno et al., 2013; Wang et al., 2012; Wheeler et al., 2006). Integration of multi-omics data can reveal patterns of molecular interactions and mechanisms that single omics data sets cannot fully capture and is thus a crucial step in the path from data to knowledge to action (Chen et al., 2012; Mayer et al., 2011; Li-Pook-Than and Snyder, 2013; Sabidó et al., 2012; Sánchez et al., 2012; Yizhak et al., 2010; Zhang et al., 2010).

The Multi-Omics Profiling Expression Database (MOPED, http://moped.proteinspire.org) was created to meet the pressing needs and demands of life sciences researchers for a freely

[1]Bioinformatics and High-Throughput Analysis Laboratory, Center for Developmental Therapeutics, Seattle Children's Research Institute, Seattle, Washington.
[2]High-throughput Analysis Core, Seattle Children's Research Institute, Seattle, Washington.
[3]Predictive Analytics, Seattle Children's, Seattle, Washington.
[4]Data-Enabled Life Sciences Alliance (DELSA Global), Seattle, Washington.
[5]Departments of Biomedical Informatics and Medical Education and Pediatrics, University of Washington, Seattle, Washington.

available public domain database of preprocessed expression information to complement already available data resources (Higdon et al., 2014; Kolker et al., 2012a). MOPED version 2.5 is a multi-omics resource that includes consistently processed proteomics and transcriptomics expression information. With each release, MOPED has provided more data, more visualization tools, and further improvements to the user interface. MOPED encourages researchers to submit raw or processed omics (e.g., transcriptomics, proteomics, metabolomics, etc.) data. The MOPED team will process the raw data and make it available to the user either in public or private MOPED. As such, MOPED provides a platform for data exploration by researchers, collaborators, or reviewers, helps to fulfill publication data submission requirements, and facilitates data sharing with the scientific community.

Launched initially as a proteomics database, MOPED 2.5 has now added gene expression data. As a proteomics resource, MOPED's users were from over 90 countries in 2013. The multi-omics collection of data will allow researchers to capitalize on the strengths across omics to enable more powerful analyses and complex hypothesis testing. With the addition of more than 150 gene relative expression experiments, MOPED is a paradigm shift away from isolated research silos toward community-wide, data-driven biological discovery. Building on the best practices used in different fields, MOPED integrates heterogeneous data into a unified public resource. The integration of multi-omics data can be essential for scientific discovery (Efron and Tibshirani, 2007; Huang, 2014; Huang et al., 2014; Olex et al., 2014), thus by providing a platform for multi-omics data, MOPED can act as a launching point for scientific discoveries.

Data Sources

MOPED 2.5 encompasses approximately 5 million transcriptomic and proteomic expression records from over 250 experiments covering four organisms: human, mouse, worm, and yeast. These expression records come from almost 200 tissues and include nearly 390 conditions.

MOPED added transcriptomics relative expression data that compares the expression of mRNAs in different conditions or tissues. The gene expression data comes from preprocessed Gene Expression Omnibus (GEO) data (Barrett et al., 2012; Edgar et al., 2002). The GEO microarray experiments were downloaded, reviewed, and analyzed using LIMMA package in R (Smyth, 2005). Results were filtered to only include gene expression for the primary organism.

Along with MOPED's expansion into a multi-omics database, the original proteomics data has continued to grow. This latest release increases the number of proteomics records to more than 500,000, a 2.5-fold increase. MOPED's protein expression sources include PeptideAtlas, PRIDE, ProteomicsDB, and collaborators (Desiere, 2006; Vizcaíno et al., 2013, https://www.proteomicsdb.org).

To provide robust and standardized analysis of proteomics data, protein identification and expression analysis are carried out using SPIRE, a proteomics analysis pipeline that integrates search engines X!Tandem and OMSSA with peptide identification models, IPM (Integrated Protein Model) and relative expression analysis (Craig et al., 2004; Geer et al., 2004; Hather et al., 2010; Higdon et al., 2008, 2011; Kolker et al., 2011, 2012b). SPIRE can directly generate protein absolute and relative expression data in a format that can be uploaded to MOPED. The absolute expression table displays concentrations, based on spectral counts and known tissue protein concentrations, in ppm, ng/mL, and nM. MOPED uniquely calculates the protein concentration dependent on the source tissue, allowing for more accurate comparisons (Higdon et al., 2014).

Protein relative expression experiments are also displayed in MOPED. These allow users to examine expression differences within comparative experiments, such as comparisons of disease state and nondisease state samples. MOPED reports expression ratios, *p*-values, and false discovery rates that are calculated based on pair-wise comparisons made using SPIRE (Higdon et al., 2014; Kolker et al., 2011, 2012b). Comparisons within an experiment can be more accurate because of consistent experimental design and therefore provide further insight into complex biological functions. Differences in protein expression may reveal the protein(s) involved in cellular responses to the condition being examined.

Using these consistent analysis methods, MOPED can also process raw omics (transcriptomics, proteomics, metabolomics, etc.) data submitted by researchers. Expression data will then be presented in a summarized form within the MOPED data interface. Researchers can choose to keep data private, which allows collaborators or reviewers to explore the data, or make the data public.

For transcriptomics and proteomics data, the MOPED team reviews experimental design, analysis methods, and results. The experiments are reviewed through metadata provided at source sites and published articles about the datasets. The DELSA proposed metadata checklist helps facilitate checking of data sources, analyses, and results (Ioannidis and Khoury, 2011; Kolker, 2013; Kolker et al., 2012c; Ozdemir et al., 2011a; 2013a, 2013b). This manual curation process increases the data quality in MOPED.

MOPED Data Interface

The data can be accessed through three ''tabs'' in MOPED: Protein Absolute Expression, Protein Relative Expression, and Gene Relative Expression. The Protein Absolute Expression tab enables the user to examine protein concentrations within and across experiments. Users can explore ratios of protein or transcript concentrations in comparative experiments within the Protein Relative Expression and Gene Relative Expression tabs.

In each of the three tabs, users can search expression data by gene symbol, UniProt protein ID, localization, condition, tissue, experiment, or keyword (Fig. 1). Terms can also be combined and filters set for more advanced searches. An example of the Gene Relative Expression Results table's content is found in Figure 1. For efficient searching, MOPED utilizes Lucene for full text indexing and AspectJ for tracking and optimization (Apache Foundation, http://lucene.apache.org/; Eclipse Foundation, http://eclipse.org/aspectj/).

Concise background information on specific proteins and genes can be found on the Protein and Gene Summary pages with corresponding proteins and genes linked, allowing exploration of multi-omics expression. The pages display information such as annotation, chromosome location,

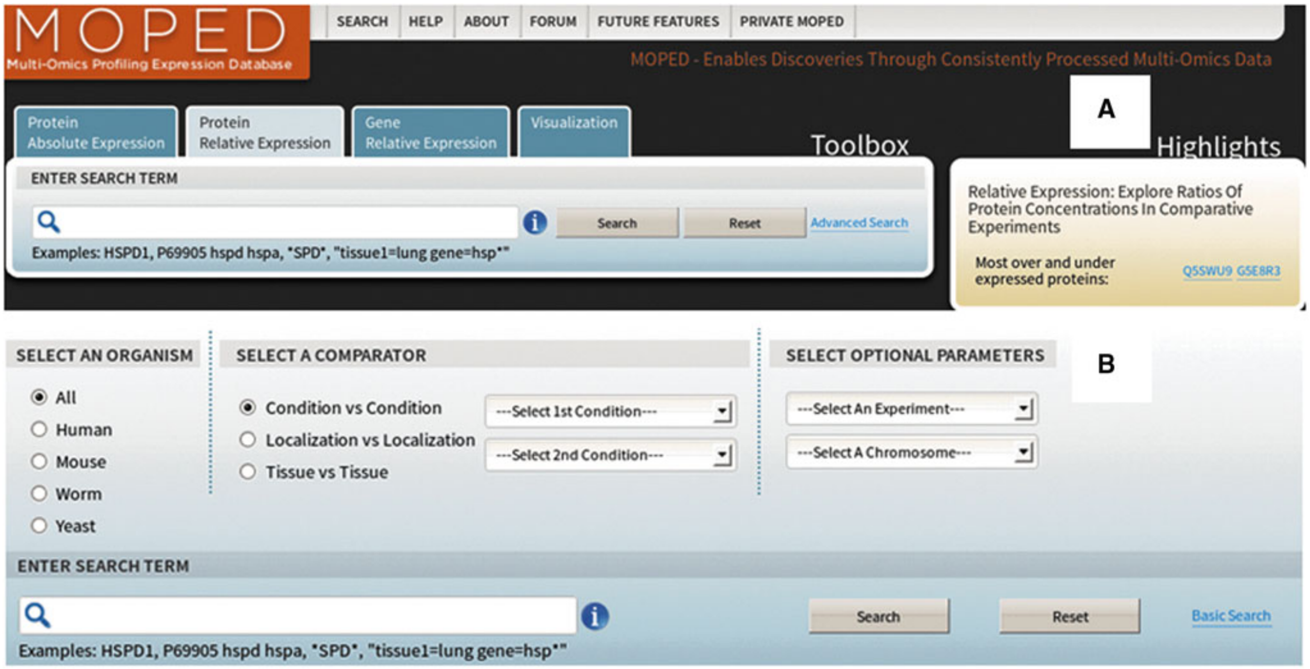

MOPED
Multi-Omics Profiling Expression Database
SEARCH
HELP
ABOUT
FORUM
FUTURE FEATURES
PRIVATE MOPED
MOPED - Enables Discoveries Through Consistently Processed Multi-Omics Data
Protein Absolute Expression
Protein Relative Expression
Gene Relative Expression
Visualization
A
Toolbox
Highlights
ENTER SEARCH TERM
Search
Reset
Advanced Search
Examples: HSPD1, P69905 hspd hspa, *SPD*, "tissue1=lung gene=hsp*"
Relative Expression: Explore Ratios Of Protein Concentrations In Comparative Experiments
Most over and under expressed proteins:
Q5SWU9 G5E8R3
SELECT AN ORGANISM
● All
○ Human
○ Mouse
○ Worm
○ Yeast
SELECT A COMPARATOR
● Condition vs Condition
○ Localization vs Localization
○ Tissue vs Tissue
---Select 1st Condition---
---Select 2nd Condition---
SELECT OPTIONAL PARAMETERS
---Select An Experiment---
---Select A Chromosome---
B
ENTER SEARCH TERM
Search
Reset
Basic Search
Examples: HSPD1, P69905 hspd hspa, *SPD*, "tissue1=lung gene=hsp*"

FIG. 1. Search options for Relative Expression tab. Both basic **(A)** and advanced **(B)** search are available for Absolute and Relative Expression tabs. Users can search by protein ID, gene symbol, tissue, condition, and/or keyword.

expression information, links to pathways from Reactome, PANTHER, and BioCyc, and external links such as GeneCards, NCBI Entrez, and the Protein Data Bank (Ashburner et al., 2000; Benson et al., 2013; Berman et al., 2000; Bult et al., 2007; Caspi et al., 2010, 2014; Cherry et al., 2012; Croft et al., 2014; Donna Maglott et al., 2013; Flicek et al., 2014; Gray et al., 2013; Kanehisa and Goto, 2000; Kanehisa et al., 2014; Mi et al., 2013; Milacic et al., 2012; Pruitt et al., 2014; Rebhan et al., 1997; Stelzer et al., 2011; The UniProt Consortium, 2014; Yook et al., 2012).

Researchers can explore protein and gene expression data in MOPED, along with complementary information about chromosome, pathway, and annotation. Data-driven science tends to begin with data exploration in order to generate hypotheses (Ozdemir et al., 2011b). By providing consistently processed expression data, MOPED becomes a platform for data exploration and can accelerate hypothesis generation. As a resource, MOPED can be used for data validation and exploration (Chen and Penning, 2014; Staneva et al., 2013; Starkey and Tilton, 2012; Williams et al., 2014). Whether starting from a protein or gene ID, an experimental condition, or just a hunch, MOPED offers a way to explore data, share knowledge and find answers.

Protein expression data

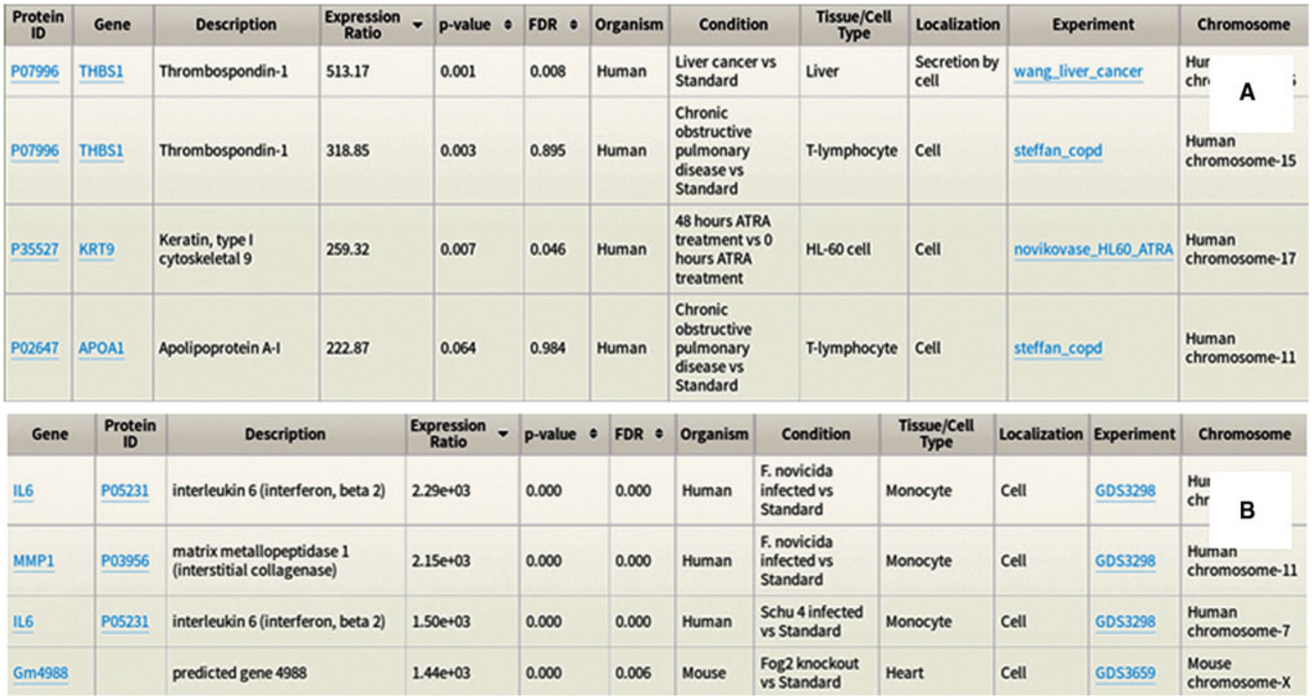
Searches within the Protein Absolute Expression tab and Protein Relative Expression tab return both expression data and experiment summaries. For Protein Absolute Expression, the default result expression table includes gene, protein name, concentration (ppm), organism, condition, tissue, and experiment. The extended view has additional fields including concentration (ng/mL, nM), false discovery rate, spectral counts, unique peptides, sequence coverage, and chromosome.

Protein Relative Expression Summary pages display concentration ratios in comparative experiments. This enables users to explore differences in expression across a wide range of condition pairs (e.g., trauma vs. standard, cystic fibrosis vs. standard, HIV positive vs. HIV resistant, and so on). The default view for relative expression summaries includes expression ratios with the corresponding *p*-value and the false discovery rate available under the extended view (Fig. 2A) (Higdon et al., 2014).

Protein Details pages can be accessed by clicking on protein IDs in MOPED. The built-in visualizations for concentrations and relative expression ratios allows for at-a-glance comparisons. Visualizations for that specific protein can also be accessed through the Protein Details page. Absolute and Relative bar charts and matrices display protein expression across tissues, conditions, localization, and experiments (Fig. 3B). MOPED also links to a number of external resources such as GeneCards for additional information (Ashburner et al., 2000; Benson et al., 2013; Berman et al., 2000; Bult et al., 2007; Caspi et al., 2010, 2014; Cherry et al., 2012; Croft et al., 2014; Donna Maglott et al., 2013; Flicek et al., 2014; Gray et al., 2013; Kanehisa et al., 2014; Kanehisa and Goto, 2000; Mi et al., 2013; Milacic et al., 2012; Pruitt et al., 2014; Rebhan et al., 1997; Stelzer et al., 2011; The UniProt Consortium, 2014; Yook et al., 2012). As a complementary resource, GeneCards offers in-depth gene information. GeneCards integrates MOPED's absolute protein expression data to create their own protein expression figure that can be accessed through a link on the MOPED Protein Details page (Rebhan et al., 1997; Stelzer et al., 2011).

Gene relative expression data

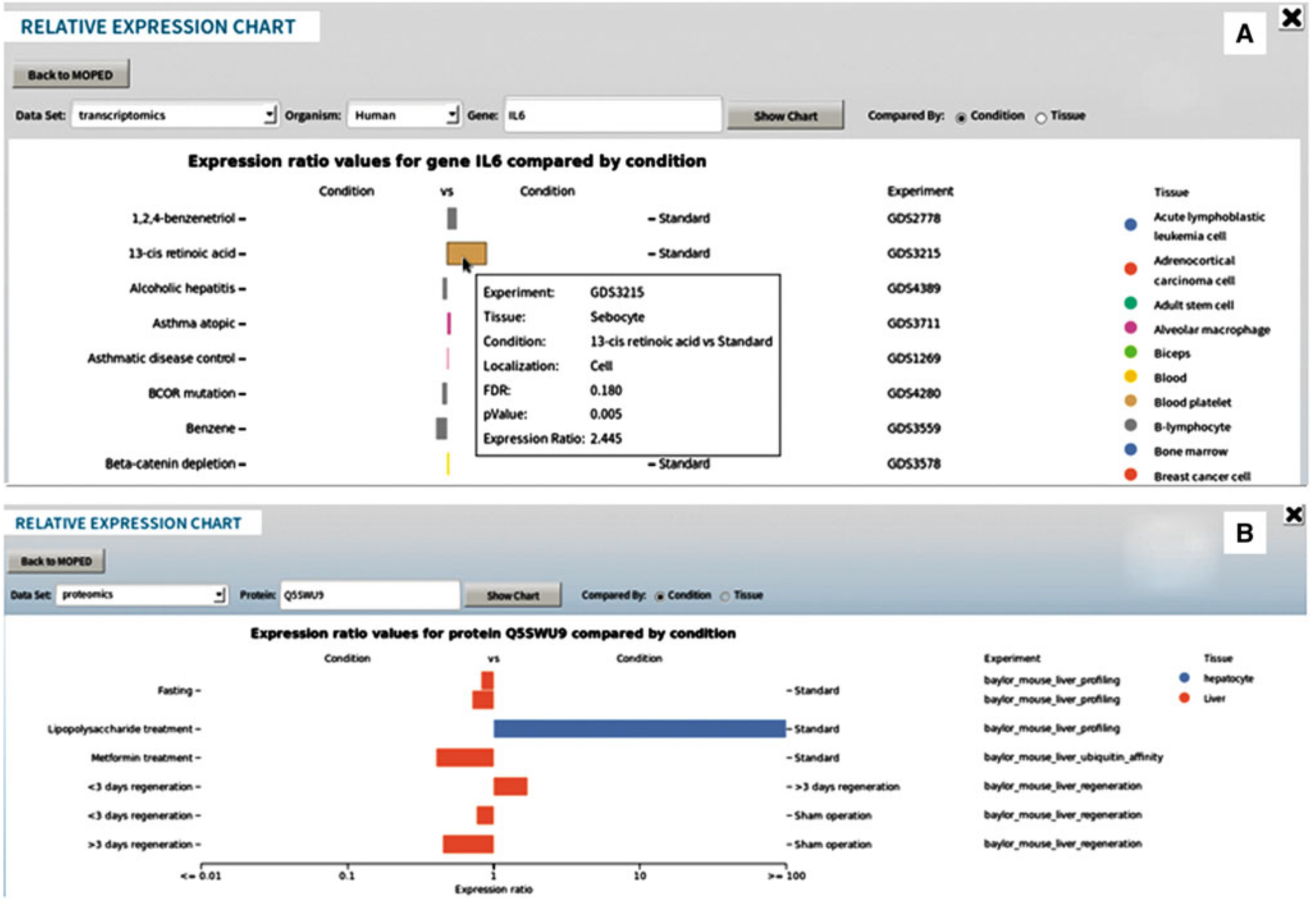
As with protein data, the Gene Relative Expression tab offers an intuitive query interface. Searching by keyword,



Protein ID	Gene	Description	Expression Ratio	p-value	FDR	Organism	Condition	Tissue/Cell Type	Localization	Experiment	Chromosome
P07996	THBS1	Thrombospondin-1	513.17	0.001	0.008	Human	Liver cancer vs Standard	Liver	Secretion by cell	wang_liver_cancer	Hur chr
P07996	THBS1	Thrombospondin-1	318.85	0.003	0.895	Human	Chronic obstructive pulmonary disease vs Standard	T-lymphocyte	Cell	steffan_copd	Human chromosome-15
P35527	KRT9	Keratin, type I cytoskeletal 9	259.32	0.007	0.046	Human	48 hours ATRA treatment vs 0 hours ATRA treatment	HL-60 cell	Cell	novikovase_HL60_ATRA	Human chromosome-17
P02647	APOA1	Apolipoprotein A-I	222.87	0.064	0.984	Human	Chronic obstructive pulmonary disease vs Standard	T-lymphocyte	Cell	steffan_copd	Human chromosome-11

Gene	Protein ID	Description	Expression Ratio	p-value	FDR	Organism	Condition	Tissue/Cell Type	Localization	Experiment	Chromosome
IL6	P05231	interleukin 6 (interferon, beta 2)	2.29e+03	0.000	0.000	Human	F. novicida infected vs Standard	Monocyte	Cell	GDS3298	Hu chr
MMP1	P03956	matrix metallopeptidase 1 (interstitial collagenase)	2.15e+03	0.000	0.000	Human	F. novicida infected vs Standard	Monocyte	Cell	GDS3298	Human chromosome-11
IL6	P05231	interleukin 6 (interferon, beta 2)	1.50e+03	0.000	0.000	Human	Schu 4 infected vs Standard	Monocyte	Cell	GDS3298	Human chromosome-7
Gm4988		predicted gene 4988	1.44e+03	0.000	0.006	Mouse	Fog2 knockout vs Standard	Heart	Cell	GDS3659	Mouse chromosome-X

FIG. 2. Search results for Protein **(A)** and Gene **(B)** Relative Expression tab in expanded view. The Protein Absolute Expression search results include Concentration in ppm, instead of Expression Ratio.



Data Set: transcriptomics Organism: Human Gene: IL6 Compared By: ● Condition ○ Tissue

Experiment: GDS3215
Tissue: Sebocyte
Condition: 13-cis retinoic acid vs Standard
Localization: Cell
FDR: 0.180
pValue: 0.005
Expression Ratio: 2.445

Data Set: proteomics Protein: Q5SWU9 Compared By: ● Condition ○ Tissue

FIG. 3. Relative Expression bar chart of genes **(A)** and proteins **(B)** displays relative expression of proteins and genes across conditions and tissues.

batch, wild card, and advanced search options are available. By default, results display gene symbol, protein ID, expression ratio, organism, condition, tissue/cell type, and experiment. The expanded view includes description, *p*-value, FDR, localization, and chromosome. Results are sortable by expression ratio, *p*-value, and FDR (Fig. 2B).

The Gene Details page offers concise gene and expression information. Similar to the Protein Details page, the Gene Details page includes a link to the corresponding protein, chromosome mapping, external links (including GeneCards, NCBI Entrez ID, etc.), gene expression visualizations, and relative expression data (Fig. 4) (Ashburner et al., 2000; Benson et al., 2013; Donna Maglott et al., 2013; Flicek et al., 2014; Gray et al., 2013; Rebhan et al., 1997; Stelzer et al., 2011; Yook et al., 2012). A visual representation of relative expression ratios is integrated into the gene expression table, allowing for simple comparisons. Links between proteins and genes provide a multi-omics understanding of protein and gene expression.

MOPED aggregates different data sources including gene expression datasets, chromosome mappings, and external IDs to present summarized information to the user. Two of the major challenges of data-driven scientific research are the cleaning of data and the modification of analysis tools (Barga et al., 2011). MOPED's data integration will help scientists explore accurate and consistent gene and protein information without the need for in-house expertise in data management and analysis tool development.

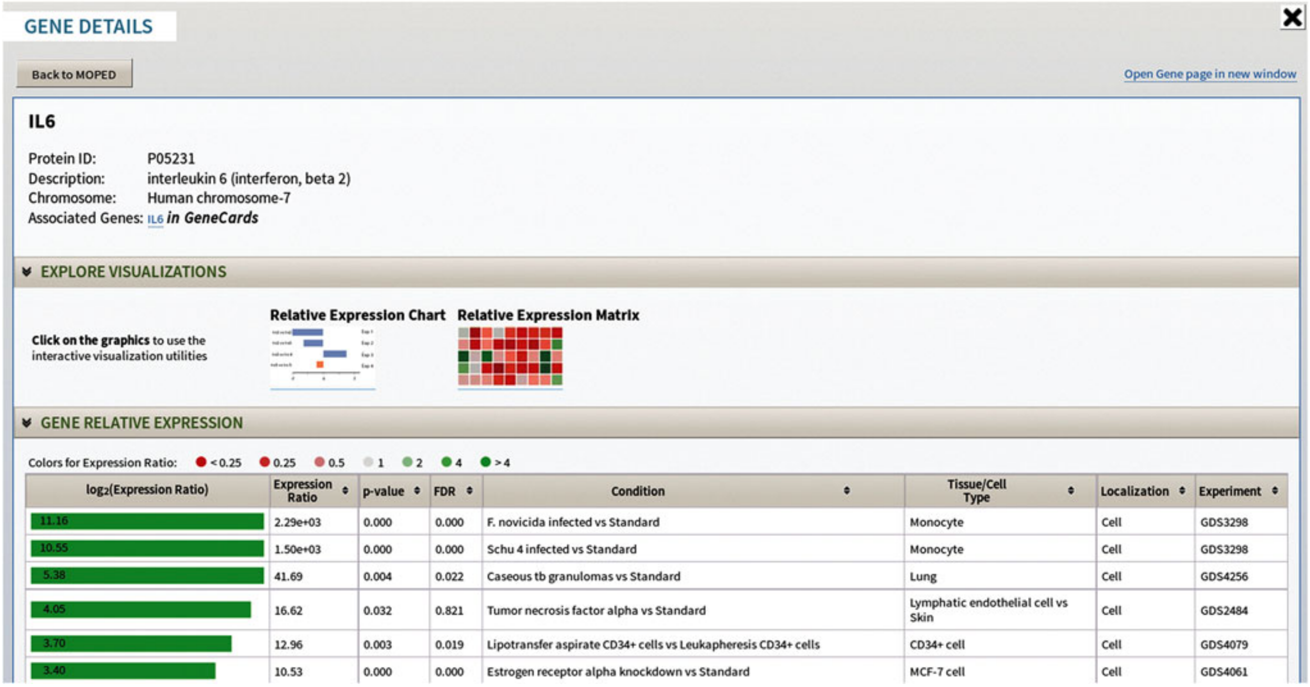
In MOPED 2.5, both protein and gene search results display the corresponding chromosomal location. The advanced search option enables users to retrieve the chromosome-specific data. This feature is aligned with the goals of the Chromosome Centric Human Proteome Project (C-HPP) as it enables researchers to map proteomics expression data to chromosome location (http://www.thehpp.org) (Fig. 5) (Paik et al., 2012).

Experiment metadata

In order to support the goal of reproducible science as pioneered by *Nature*, MOPED 2.5 now provides detailed experimental metadata, accessible through the Experiment Summary pages (2013b). Endorsed by DELSA Global, the metadata checklist provides information about experimental design, instrument details, sample preparation, data processing, and analysis (Ioannidis and Khoury, 2011; Kolker, 2013; Kolker et al., 2012c; Ozdemir et al., 2011a; 2013a, 2013b). Comprehensive metadata enable researchers to repeat and validate experiments (Ioannidis and Khoury, 2011). By providing metadata, MOPED allows researchers to assess the relevance and usefulness of a given dataset. In addition, the checklists will allow the user to more easily evaluate and integrate diverse data types.

Visualization tools

MOPED 2.5 offers a Visualization tab, which enables graphical exploration of data. Protein absolute expression data in MOPED 2.5 can be seen at a glance on a *chord chart* relating tissues, conditions, and localizations. *Expression bar charts* display protein or gene expression across tissues, conditions, and experiments (Fig. 2). The data are grouped by tissue, color-coded by condition, and searchable by ID. Built-in hovering displays detailed experimental information. The bar charts are searchable by protein or gene. *Expression matrices* display the expression data of one protein or one gene and up to 10 neighboring proteins or genes to


GENE DETAILS
Back to MOPED
Open Gene page in new window
EXPLORE VISUALIZATIONS
Relative Expression Chart
Relative Expression Matrix
GENE RELATIVE EXPRESSION

IL6

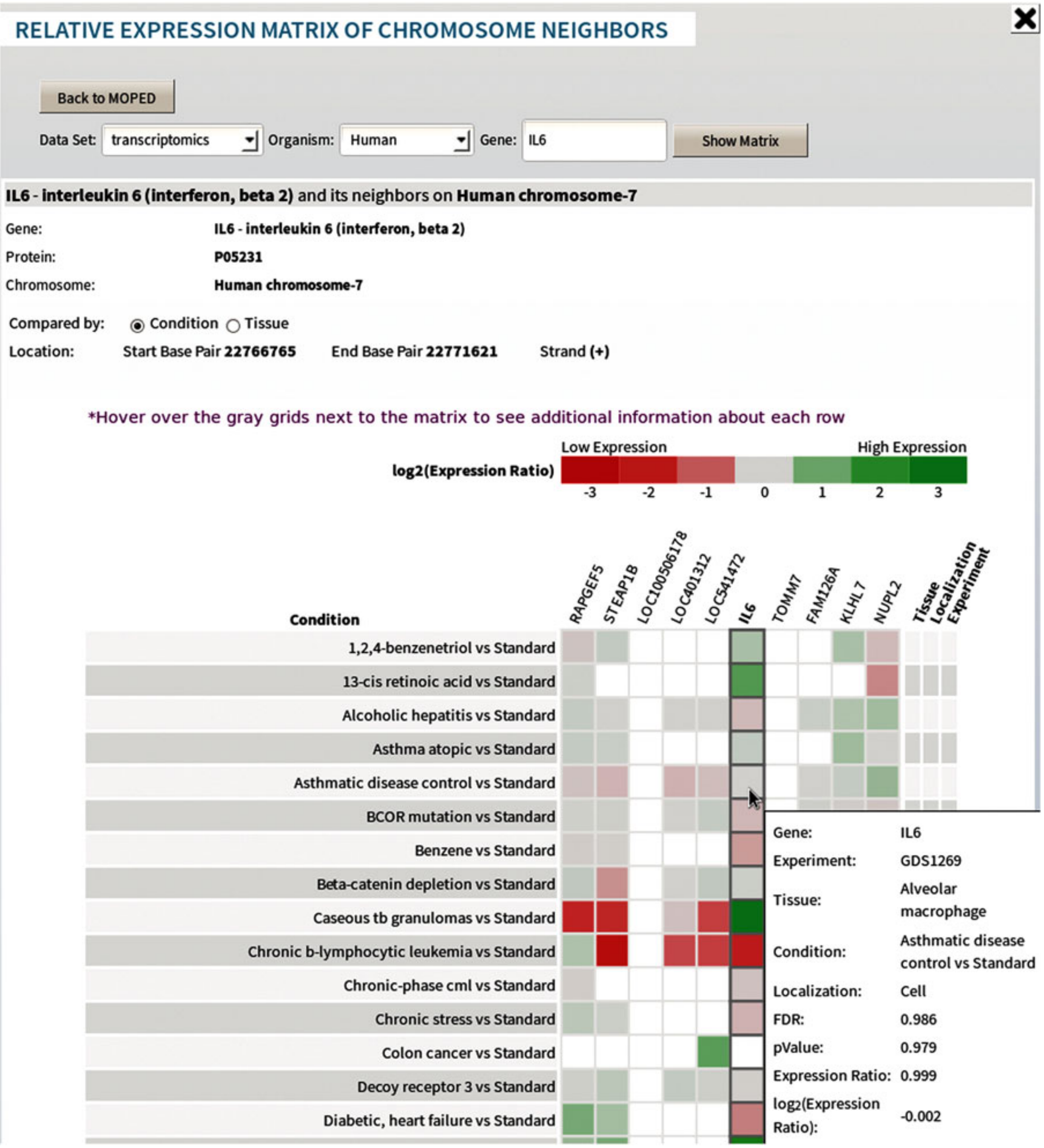
Protein ID: P05231
Description: interleukin 6 (interferon, beta 2)
Chromosome: Human chromosome-7
Associated Genes: IL6 *in GeneCards*

Click on the graphics to use the interactive visualization utilities

Colors for Expression Ratio: ● < 0.25 ● 0.25 ● 0.5 ● 1 ● 2 ● 4 ● > 4

log2(Expression Ratio)	Expression Ratio	p-value	FDR	Condition	Tissue/Cell Type	Localization	Experiment
11.16	2.29e+03	0.000	0.000	F. novicida infected vs Standard	Monocyte	Cell	GDS3298
10.55	1.50e+03	0.000	0.000	Schu 4 infected vs Standard	Monocyte	Cell	GDS3298
5.38	41.69	0.004	0.022	Caseous tb granulomas vs Standard	Lung	Cell	GDS4256
4.05	16.62	0.032	0.821	Tumor necrosis factor alpha vs Standard	Lymphatic endothelial cell vs Skin	Cell	GDS2484
3.70	12.96	0.003	0.019	Lipotransfer aspirate CD34+ cells vs Leukapheresis CD34+ cells	CD34+ cell	Cell	GDS4079
3.40	10.53	0.000	0.000	Estrogen receptor alpha knockdown vs Standard	MCF-7 cell	Cell	GDS4061

FIG. 4. The Gene Details page displays gene specific information. The expression ratio bar chart is color-coded based on expression, providing at-a-glance comparison of relative gene expression.



RELATIVE EXPRESSION MATRIX OF CHROMOSOME NEIGHBORS

Data Set: transcriptomics Organism: Human Gene: IL6

IL6 - interleukin 6 (interferon, beta 2) and its neighbors on **Human chromosome-7**

Gene:	**IL6 - interleukin 6 (interferon, beta 2)**
Protein:	**P05231**
Chromosome:	**Human chromosome-7**

Compared by: ● Condition ○ Tissue

Location: Start Base Pair **22766765** End Base Pair **22771621** Strand **(+)**

*Hover over the gray grids next to the matrix to see additional information about each row

Condition

1,2,4-benzenetriol vs Standard
13-cis retinoic acid vs Standard
Alcoholic hepatitis vs Standard
Asthma atopic vs Standard
Asthmatic disease control vs Standard
BCOR mutation vs Standard
Benzene vs Standard
Beta-catenin depletion vs Standard
Caseous tb granulomas vs Standard
Chronic b-lymphocytic leukemia vs Standard
Chronic-phase cml vs Standard
Chronic stress vs Standard
Colon cancer vs Standard
Decoy receptor 3 vs Standard
Diabetic, heart failure vs Standard

Gene:	IL6
Experiment:	GDS1269
Tissue:	Alveolar macrophage
Condition:	Asthmatic disease control vs Standard
Localization:	Cell
FDR:	0.986
pValue:	0.979
Expression Ratio:	0.999
log2(Expression Ratio):	-0.002

FIG. 5. Relative Expression matrix of neighboring genes along the chromosome. Expression ratios are color-coded by expression and from comparative experiments. Proteins corresponding to neighboring genes along the chromosome can also be displayed in this form.

build a chromosomal neighborhood view of expression. The expression values are color-coded and listed by tissues, conditions, localizations, and experiments. By facilitating comparisons of protein or gene expression along the chromosome, the expression matrices further advance the efforts of the Chromosome Centric Human Proteome Project (C-HPP) (http://www.thehpp.org) (Fig. 5) (Paik et al., 2012).

Community Involvement: MOPED Forum

Continual improvement and expansion of MOPED are driven by the needs and feedback of the Life Sciences community (Higdon et al., 2014; Kolker et al., 2012c; Ozdemir et al., 2011a). The database was originally created due to survey responses in which researchers stated that a proteomics resource that builds upon already available data repositories would be extremely helpful (Higdon et al., 2013, 2014).

To enable more efficient communication with users, MOPED now includes a public forum that can be accessed through the Forum link on the home page. The science community can post questions that will be answered by the MOPED team, suggest future features for development, and get help with any usage issues. Students are encouraged to participate with an ‘‘Ask an Omics Scientist’’ section providing a discussion opportunity between future and current omics scientists.

Future Features

To give further insight into molecular mechanisms through multi-omics expression data analysis, MOPED plans to integrate pathways and disease information with the currently available data. Comparing the expression of genes and proteins along pathways will uncover intricacies of molecular interactions linked to disease states, leading to further understanding of the disease and ultimately, improved treatments. Using already developed pathway analysis tools, for example DEAP, the expression of proteins and genes can be analyzed along a pathway (Haynes et al., 2013; Subramanian et al., 2005). In addition, data for two other omics, metabolomics and lipidomics, are being assessed for incorporation into MOPED.

Data Submission

Raw or processed omics data (transcriptomics, proteomics, metabolomics, etc.) can be submitted to MOPED by either on-line upload or mailing a hard drive. Researchers, who wish to upload data to public or private MOPED, may contact us on the MOPED Forum (moped-forum.proteinspire.org). Experiment metadata checklists should accompany data and provide accurate experimental and analytical methods in order to increase reproducibility (Ioannidis and Khoury, 2011; Kolker, 2013; Kolker et al., 2012c; Ozdemir et al., 2011a; 2013a, 2013b). Researchers can submit metadata checklists as data publications to journals, for example, OMICS (Snyder et al., 2014). Data can also be uploaded to private MOPED where it can be shared with collaborators and reviewers.

Conclusion

The biological functions of organisms depend on complex and highly interactive systems of biomolecules including RNA, proteins, metabolites, and lipids. These biomolecules are characterized by high-throughput multi-omics data from transcriptomics, proteomics, metabolomics, and lipidomics experiments. Data-enabled biological discoveries require high-throughput data to be integrated and analyzed jointly across multi-omics experiments, yet developing a useful integrated resource is challenging due to the scale of data and complexity of the technologies, formats, ontologies, and methodologies. The collective efforts of multiple disciplines must be used to confront these challenges. MOPED 2.5 has taken a significant step towards overcoming these challenges by becoming a multi-omics database with consistently processed transcriptomics and proteomics data from over 250 experiments.

MOPED 2.5 is a pivotal public database and interdisciplinary knowledge platform for 21st century integrative biology applications from lab to society. Through integrated multi-omics data, MOPED can serve as a platform for multi-omics life science discoveries.

Acknowledgments

We thank Maggie Lackey for her critical reading. Research reported in this study was supported by the National Science Foundation under the Division of Biological Infrastructure award 0969929, National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health under awards U01DK089571 and U01DK072473, and awards from Seattle Children's Research Institute, The Robert B. McMillen Foundation, and The Gordon and Betty Moore Foundation awarded to E.K.. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Science Foundation, National Institute of Diabetes and Digestive and Kidney Metabolites, Diseases of the National Institutes of Health, Seattle Children's Research Institute, The Robert B. McMillen Foundation, and The Gordon and Betty Moore Foundation.

Author Disclosure Statement

The authors declare no competing financial interests exist.

References

Ashburner M, Ball CA, Blake JA, et al. (2000). Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. Nature Genetics 25, 25–29.

Barga R, Howe B, Beck D, et al. (2011). Bioinformatics and data-intensive scientific discovery in the beginning of the 21st century. Omics 15, 199–201.

Barrett T, Wilhite SE, Ledoux P, et al. (2012). NCBI GEO: Archive for functional genomics data sets—update. Nucleic Acids Res 41, D991–D995.

Benson DA, Cavanaugh M, Clark K, et al. (2013). GenBank. Nucleic Acids Res 41, D36–42.

Berman HM, Westbrook J, Feng Z, et al. (2000). The Protein Data Bank. Nucleic Acids Res 28, 235–242.

Bult CJ, Eppig JT, Kadin JA, Richardson JE, Blake JA, and the Mouse Genome Database Group. (2007). The Mouse Genome Database (MGD): Mouse biology and model systems. Nucleic Acids Res 36, D724–D728.

Caspi R, Altman T, Dale JM, et al. (2010). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res 38, D473–479.

Caspi R, Altman T, Billington R, et al. (2014). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Res 42, D459–D471.

Chen M, and Penning TM. (2014). 5β-Reduced steroids and human Δ(4)-3-ketosteroid 5β-reductase (AKR1D1). Steroids 83, 17–26.

Chen R, Mias GI, Li-Pook-Than J, et al. (2012). Personal omics profiling reveals dynamic molecular and medical phenotypes. Cell 148, 1293–1307.

Cherry JM, Hong EL, Amundsen C, et al. (2012). Saccharomyces Genome Database: The genomics resource of budding yeast. Nucleic Acids Res 40, D700–705.

Craig R, Beavis RC. (2004). TANDEM: matching proteins with tandem mass spectra. Bioinformatics 20, 1466-1467.

Croft D, Mundo AF, Haw R, et al. (2014). The Reactome pathway knowledgebase. Nucleic Acids Res 42, D472–477.

Desiere F. (2006). The PeptideAtlas project. Nucleic Acids Res 34, D655–D658.

Edgar R, Domrachev M, and Lash AE. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 30, 207–210.

Efron, B and Tibshirani R (2007). On testing the significance of sets of genes. The Annals of Applied Statistics 1, 107–129.

Flicek P, Amode MR, Barrell D, et al. (2014). Ensembl 2014. Nucleic Acids Res 42, D749–D755.

Geer LY, Markey SP, Kowalak JA, et al. (2004). Open mass spectrometry search algorithm. J Proteome Res 3, 958–964.

Gray KA, Daugherty LC, Gordon SM, Seal RL, Wright MW, and Bruford EA. (2013). Genenames.org: The HGNC resources in 2013. Nucleic Acids Res 41, D545–552.

Hather G, Higdon R, Bauman A, von Haller PD, and Kolker E. (2010). Estimating false discovery rates for peptide and protein identification using randomized databases. Proteomics 10, 2369–2376.

Haynes WA, Higdon R, Stanberry L, Collins D, and Kolker E. (2013). Differential expression analysis for pathways. PLoS Computational Biology 9, e1002967.

Higdon R, van Belle G, and Kolker E. (2008). A note on the false discovery rate and inconsistent comparisons between experiments. Bioinformatics 24, 1225–1228.

Higdon R, Reiter L, Hather G, et al. (2011). IPM: An integrated protein model for false discovery rate estimation and identification in high-throughput proteomics. J Proteomics 75, 116–121.

Higdon R, Haynes W, Stanberry L, et al. (2013). Unraveling the complexities of life sciences data. Big Data 1, 42–50.

Higdon R, Stewart E, Stanberry L, et al. (2014). MOPED enables discoveries through consistently processed proteomics data. J Proteome Res 13, 107–113.

Huang Y-T, VanderWeele TJ, Lin X. (2014). Joint analysis of SNP and gene expression data in genetic association studies of complex diseases. The Annals of Applied Statistics 8, 352–376.

Huang Y-T. (2014). Integrative modeling of multiple genomic data from different types of genetic association studies. Biostatistics (Oxford, England).

Ioannidis JPA, and Khoury MJ. (2011). Improving validation practices in ''omics'' research. Science 334, 1230–1232.

Kanehisa M, and Goto S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28, 27–30.

Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, and Tanabe M. (2014). Data, information, knowledge and principle: Back to metabolism in KEGG. Nucleic Acids Res 42, D199–205.

Khatri P, Sirota M, and Butte AJ. (2012). Ten years of pathway analysis: Current approaches and outstanding challenges. PLoS Comput Biol 8, e1002375.

Kolker E, Altintas I, Bourne P, et al. (2013). Reproducibility: In praise of open research measures. Nature 498, 170–170.

Kolker E, Higdon R, Welch D, et al. (2011). SPIRE: Systematic protein investigative research environment. J Proteomics 75, 122–126.

Kolker E, Higdon R, Haynes W, et al. (2012a). MOPED: Model Organism Protein Expression Database. Nucleic Acids Res 40, D1093–1099.

Kolker E, Higdon R, Welch D, et al. (2012b). Corrigendum to ''SPIRE: Systematic Protein Investigative Research Environment'' [J. Proteomics 75 (1) (2011) 122–126]. J Proteomics 75, 3789.

Kolker E, Stewart E, and Özdemir V. (2012c). DELSA Global for ''Big Data'' and the bioeconomy: Catalyzing collective innovation. Indust Biotechnol 8, 176–178.

Maglott D. (2004). Entrez Gene: Gene-centered information at NCBI. Nucleic Acids Res 33, D54–D58.

Maglott D, Pruitt K, Tatusova T, and Terence M. (2013). Gene. In: *The NCBI Handbook*, (Bethesda, MD: National Center for Biotechnology Information (US)).

Mayer C-D, Lorent J, and Horgan GW. (2011). Exploratory analysis of multiple omics datasets using the adjusted RV coefficient. Statist Applicat Genetics Mol Biol 10, Article 14.

Mi H, Muruganujan A, and Thomas PD. (2013). PANTHER in 2013: Modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Res 41, D377–386.

Milacic M, Haw R, Rothfels K, et al. (2012). Annotating cancer variants and anti-cancer therapeutics in reactome. Cancers 4, 1180–1211.

Nature's Guide for Authors (2013a). Reporting Checklist for Life Science Articles. Nature: http://www.nature.com/authors/policies/checklist.pdf

Nature's Editorial (2013b). Announcement: Reducing our irreproducibility. Nature 496, 398. http://www.nature.com/news/announcement-reducing-our-irreproducibility–1.12852

Olex AL, Turkett WH, Fetrow JS, Loeser RF (2014). Integration of the expression data with network-based analysis to identify signaling and metabolic pathways regulated during the development of osteoarthritis. Gene 542, 38–45.

Ozdemir V, Rosenblatt DS, Warnich L, et al. (2011a). Towards an ecology of collective innovation: Human Variome Project (HVP), Rare Disease Consortium for Autosomal Loci (RaDiCAL) and Data-Enabled Life Sciences Alliance (DELSA). Curr Pharmacogenom Personalized Med 9, 243–251.

Ozdemir V, Smith C, Bongiovanni K, et al. (2011b). Policy and data-intensive scientific discovery in the beginning of the 21st century. OMICS 15, 221–225.

Paik Y-K, Jeong S-K, Omenn GS, et al. (2012). The Chromosome-Centric Human Proteome Project for cataloging proteins encoded in the genome. Nature Biotechnol 30, 221–223.

Parkinson H, Sarkans U, Kolesnikov N, et al. (2011). ArrayExpress update—An archive of microarray and high-throughput sequencing-based functional genomics experiments. Nucleic Acids Res 39, D1002–1004.

Li-Pook-Than J, and Snyder M. (2013). iPOP goes the world: Integrated personalized omics profiling and the road toward improved health care. Chem Biol 20, 660–666.

Pruitt KD, Brown GR, Hiatt SM, et al. (2014). RefSeq: An update on mammalian reference sequences. Nucleic Acids Res 42, D756–763.

R Core Team. (2013). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rebhan M, Chalifa-Caspi V, Prilusky J, and Lancet D. (1997). GeneCards: Integrating information about genes, proteins and diseases. Trends Genetics 13, 163.

Sabidó E, Quehenberger O, Shen Q, et al. (2012). Targeted proteomics of the eicosanoid biosynthetic pathway completes an integrated genomics-proteomics-metabolomics picture of cellular metabolism. Mol Cell Proteomics 11, M111.014746.

Sánchez A, Fernández-Real J, Vegas E, et al. (2012). Multivariate methods for the integration and visualization of omics data. In *Bioinformatics for Personalized Medicine*, A.T. Freitas, and A. Navarro, eds. (Springer Berlin Heidelberg) pp. 29–41.

Smyth GK. (2005). Limma: linear models for microarray data. In *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, W. Huber, eds. (New York: Springer): pp. 397–420.

Snyder M, Mias G, Stanberry L, and Kolker E. (2014). Metadata checklist for the integrated personal omics study: Proteomics and metabolomics experiments. OMICS 18, 81–85 also in Big Data 1(4): 202–206.

Staneva R, Rukova B, Hadjidekova S, et al. (2013). Whole genome methylation array analysis reveals new aspects in Balkan endemic nephropathy etiology. BMC Nephrology 14, 225.

Starkey JM, and Tilton RG. (2012). Proteomics and systems biology for understanding diabetic nephropathy. J Cardiovasc Translat Res 5, 479–490.

Stelzer G, Dalah I, Stein TI, et al. (2011). In-silico human genomics with GeneCards. Human Genomics 5, 709–717.

Subramanian A, Tamayo P, Mootha VK, et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 102, 15545–15550.

The UniProt Consortium. (2014). Activities at the Universal Protein Resource (UniProt). Nucleic Acids Res 42, D191–D198.

Vizcaíno JA, Côté RG, Csordas A, et al. (2013). The PRoteomics IDEntifications (PRIDE) database and associated tools: Status in 2013. Nucleic Acids Res 41, D1063–1069.

Wang M, Weiss M, Simonovic M, et al. (2012). PaxDb, a database of protein abundance averages across all three domains of life. Mol Cell Proteomics 11, 492–500.

Wheeler DL, Barrett T, Benson DA, et al. (2006). Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 34, D173–180.

Williams BC, Filter JJ, Blake-Hodek KA, et al. (2014). Greatwall-phosphorylated Endosulfine is both an inhibitor and a substrate of PP2A-B55 heterotrimers. eLife 3, e01695.

Yizhak K, Benyamini T, Liebermeister W, Ruppin E, and Shlomi T. (2010). Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model. Bioinformatics 26, i255–i260.

Yook K, Harris TW, Bieri T, et al. (2012). WormBase 2012: More genomes, more data, new website. Nucleic Acids Res 40, D735–D741.

Zhang W, Li F, and Nie L. (2010). Integrating multiple ‘‘omics’’ analysis for microbial biology: Application and methodologies. Microbiology 156, 287–301.

Address correspondence to:
Eugene Kolker, MD
Bioinformatics and High-Throughput Analysis Laboratory
Seattle Children's Research Institute
1900 Ninth Avenue
Seattle 98101, WA

E-mail: eugene.kolker@seattlechildrens.org