RESEARCH ARTICLE

Integrated Biomedical System [version 1; peer review: 2 not approved]

Darrell O. Ricke , James Harper, Anna Shcherbina, Nelson Chiu, Tara Boettcher

MIT Lincoln Laboratory, Lexington, MA, 02420, USA

V1 First published: 08 Feb 2018, 7:162
<https://doi.org/10.12688/f1000research.13601.1>
 Latest published: 08 Feb 2018, 7:162
<https://doi.org/10.12688/f1000research.13601.1>

Abstract

Background: Capabilities for generating and storing large amounts of data relevant to individual health and performance are rapidly evolving and have the potential to accelerate progress toward quantitative and individualized understanding of many important issues in health and medicine. Recent advances in clinical and laboratory technologies provide increasingly complete and dynamic characterization of individual genomes, gene expression levels for genes, relative abundance of thousands of proteins, population levels for thousands of microbial species, quantitative imaging data, and more – all on the same individual. Personal and wearable electronic devices are increasingly enabling these same individuals to routinely and continuously capture vast amounts of quantitative data including activity, sleep, nutrition, environmental exposures, physiological signals, speech, and neurocognitive performance metrics at unprecedented temporal resolution and scales. While some of the companies offering these measurement technologies have begun to offer systems for integrating and displaying correlated individual data, these are either closed/proprietary platforms that provide limited access to sensor data or have limited scope that focus primarily on one data domain (e.g. steps/calories/activity, genetic data, etc.).

Methods: The Integrated Biomedical System is developed as a Ruby on Rails application with a relational database.

Results: Data from multiple wearable monitors for activity, sleep, and physiological measurements, phone GPS tracking, individual genomics, air quality monitoring, etc. have been integrated into the Integrated Biomedical System.

Conclusions: The Integrated Biomedical System is being developed to demonstrate an adaptable open-source tool for reducing the burden associated with integrating heterogeneous genome, interactome, and exposome data from a constantly evolving landscape of biomedical data generating technologies. The Integrated Biomedical System provides a scalable and modular framework that can be extended to include support for numerous types of analyses and applications at scales ranging from personal users, communities and groups, to potentially large populations.

Open Peer Review

Approval Status

	1	2
version 1		
08 Feb 2018	view	view

1. **Guillermo H. Lopez-Campos**, Queen's

University of Belfast, Belfast, UK

Philip Kiossoglou, The University of

Melbourne, Parkville, Australia

2. **Wolfgang Kuchinke** , Heinrich Heine

University Düsseldorf (HHU), Düsseldorf,

Germany

Any reports and responses or comments on the article can be found at the end of the article.

Keywords

genome, exposome, interactome, exposure, wearable, health tracker, fitness device, fitness tracker, sleep, heart rate

Corresponding author: Darrell O. Ricke (Darrell.Ricke@ll.mit.edu)

Author roles: **Ricke DO:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Harper J:** Conceptualization, Funding Acquisition, Investigation, Methodology, Writing – Original Draft Preparation; **Shcherbina A:** Data Curation, Investigation, Methodology, Resources, Software; **Chiu N:** Software; **Boettcher T:** Investigation, Methodology, Resources

Competing interests: No competing interests were disclosed.

Grant information: This work is sponsored by the Assistant Secretary of Defense for Research & Engineering under Air Force Contract #FA8721-05-C-002. Opinions, interpretations, recommendations and conclusions are those of the author and are not necessarily endorsed by the United States Government.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2018 Ricke DO *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Ricke DO, Harper J, Shcherbina A *et al.* **Integrated Biomedical System [version 1; peer review: 2 not approved]** F1000Research 2018, 7:162 <https://doi.org/10.12688/f1000research.13601.1>

First published: 08 Feb 2018, 7:162 <https://doi.org/10.12688/f1000research.13601.1>

Introduction

Human health and performance is understood to be affected by both nature (genome) and nurture (activities & environment). One notable example of the combined effects of genetics and the environment on health is the identification that the GRIN2A gene significantly modulates risk for developing Parkinson's disease, but only in heavy coffee-drinkers¹. This study provides proof that inclusion of quantitative measures of environmental factors can help identify important genes that would be otherwise missed in GWAS studies that ignore exposures. However, the challenges associated with designing and implementing broad quantitative studies of complex interactions at scales sufficient to achieve sufficient statistical power are considerable.

There are multiple efforts underway that are making progress toward addressing the challenges of integrating genome, interactome, and exposome² data to support focused scientific studies. The Institute of Systems Biology's Hundred Person Wellness Project³ and 100K Project⁴ are integrating genomics, monitors, and blood sampling to build on the pioneering N-of-one work conducted by Larry Smarr⁵ and Michael Snyder^{6,7} to articulate the vision and promise of predictive, preventative, personalized, and participatory (P4) medicine⁸. Orion Bionetworks⁹ is combining traits, genetics, and interactome with a focus on brain disorders. Sanchez *et al.*¹⁰ has also proposed exposome informatics integrating the genome, phenome, and exposome. Systems integrating personal sensors and exposome have been developed by Doherty & Oh¹¹ and Nieuwenhuijsen *et al.*¹². Other relevant available resources include PhysioNet¹³ and MOPED¹⁴. The Human Longevity project¹⁵ is examining genome, microbiomes, and metabolites of volunteers. Lifestyle affects human microbiomes^{16,17}. While these projects all share the common elements of longitudinal integration of heterogeneous biomedically relevant data, each either focuses on a relatively narrow set of measurements or relies on custom data storage and analysis architectures that do not provide a scalable foundation for larger-scale integration across studies to enable meta-analysis of data from multiple studies.

The Integrated Biomedical System is being developed as an open source platform for integrating genome, interactome, and exposome data that provides a unifying model to promote more open data sharing and analysis. The software architecture with multi-scale operability design intended to scale from running on a single laptop/workstation as a standalone system with an embedded private local database, to a study platform, to large-scale implementations all using standard scalable web technology stacks.

Methods

Protocol design and approvals

The Integrated Biomedical Project description and written consent form (Protocol # 1312006029) was reviewed and approved by the Massachusetts Institute of Technology (MIT) Committee on the Use of Humans as Experimental Subjects (COUHES) for the initial 20 volunteers and the expansion to 40 volunteers. COUHES is the MIT Institutional Review Board (IRB). This project used no recruitment. All volunteers learned about the project from other volunteers, typically by expressing

interest in the devices being worn. Upon expressing interest to the principal investigator, the project was fully explained and a written consent form was provided and the project explained with multiple voluntary options. The project principal investigator and co-investigator were primary points of contacts for all volunteers. All volunteers signed the written consent form approved by the MIT COUHES and provided their signed form to the project principal investigator. Volunteers have full choice of all elements of the research project that they elect to participate in or not. Volunteers may elect to have all or any subset of their data removed from the system at any time. Volunteers either elect to opt-in or opt-out of notification of any possible data abnormalities detected.

Consent

Written informed consent was obtained from all volunteers.

Genome

Extract, transform, and load (ETL) modules were developed for 23andMe SNPs¹⁸ files, SwissProt¹⁹ dat file, DrugBank²⁰ XML file, NCBI Gene²¹ gene file, PharmGKB pathways²², and Protein Data Bank (PDB) protein structures²³. After SwissProt sequences and PDB protein structures were loaded, the structure coordinates were mapped to sequence residues with the included lib/utilities/align_pdb.rb tool; this enables the visualization of residues and variants on structures. Interface modules were developed to allow individual or pooled variants to be visualized on protein structures with the integrated Jmol²⁴ structure viewer.

Interactome

Interactome data included in the pilot collection described herein includes heart rate, interbeat interval (IBI), and electrocardiogram (ECG), skin temperature, skin conductance, galvanic skin response, and respiratory rate. These aggregated data were collected by a diverse collection of commercially available wearable physiological monitoring devices. All volunteers were offered a Basis B1 watch²⁵ and Polar Loop H7 heart rate monitor²⁶. A subset of volunteers are evaluating Hildago Equivital EQ-02-SEM²⁷, Empatica E3²⁸, Mio Link²⁹, and Zephyr BioHarness 3³⁰ devices. Data logging functionality was not built in for the Polar Loop and Mio Link heart rate monitors, so these data streams were wirelessly synced and stored continuously on co-worn Actigraph Actisleep device. ETL modules were developed for Basis B1 json files³¹, Actigraph heart rate csv or dat files (including Polar Loop and Mio Link), Empatica E3 zip files, Hidalgo Equivital SEM2 persisted summary csv files, Zephyr BioHarness summary csv files, vocal recordings and associated Matlab .mat files. Data displays include Ruby gems and JavaScript plugins: Google Maps³², jQuery³³, lazy_high_charts³⁴, Highstocks³⁵, Data-Drive Documents (D3)³⁶, FullCalendar³⁷, rails3-jquery-autocomplete³⁸, and more. The graphical user interface for "Data Loading" provides the ability to download data from the Basis web site and drag and drop interfaces for easy file uploads for each of the device ETL modules.

Exposome

Activity and sleep were monitored continuously using wearable and personal electronic devices that used algorithms to process raw data provided by built-in 3-axis accelerometers. Data describing

daily nutrition, prescriptions, and over-the-counter medications were collected manually and provided by a subset of volunteers. Devices used by volunteers for continuous data collection included the Fitbit Flex, the Basis B1 watch, Actigraph ActiSleep monitors, basic Actigraph activity monitors GT3X+, Jawbone Up, and smart-phone apps including MyTracks, and Sleep Cycle. ETL modules were developed for Fitbit csv files, Jawbone csv files, Actigraph³⁹ sleep csv files, MyTracks app⁴⁰ csv files, and Sleep Cycle app⁴¹ csv files. Additionally to demonstrate the ability to integrate other publicly available data, modules were developed for integration of EPA AirData (daily and hourly csv files⁴²), and foods⁴³. Graphical user interfaces were developed for entering activities, events, meals, drinks, prescriptions, and over-the-counter medicines. Multiple volunteers submitted oral swab samples for metagenomics sequence analysis when sick (cued data collection).

Integrated Biomedical System (iBio)

The Integrated Biomedical System was developed on the Ruby on Rails⁴⁴ platform with Ruby gems and JavaScript plugins. The Rails platform supports multiple SQL relational databases including MySQL and no SQL databases such as Mongo DB. MySQL, Oracle, Mongo DB, etc. all scale to over a billion records in a single table. The underlying architecture and approach can be extended to handle a variety of additional data sources. To facilitate data exchange between sites, global unique identifiers (guids) are used. The Integrated Biomedical System and Rails can be installed on computers ranging from stand-alone on a laptop/desktop to servers running Windows OS, Mac OS, Linux, or Unix. Individuals can install and run this system for personal use without needing to set up a web service; to facilitate this the default configuration uses the Sqlite3 database, which installs with the Rails setup. Switching to MySQL or Oracle requires database software installation and a 5-line update to the Rails database.yml configuration file with updated database instance details. To facilitate bulk loading of large numbers of data files, command line interfaces for each ETL module are included in the app/utilities folder.

Implementation

The Integrated Biomedical System (version 1.0) is developed as a Ruby on Rails (versions 3 & 4) application. Current JavaScript libraries and versions are included in the Ruby on Rails Gemfile with the inclusion of JQuery, D3, FullCalendar, Highcharts, JSmol²⁴ PDB structure viewer, and more. The Integrated Biomedical System can be optionally configured as a web site with Apache httpd web server plus Passenger (Phusion). The database schema is available in a MySQL Workbench schema in the docs folder for the application. The Integrated Biomedical System has been tested with both Sqlite3 and MySQL relational databases; it should work with most if not all Rails supported databases. The application Readme and GitHub site (<https://github.com/doricke/ibio>) list the 10 standard Rails application setup steps to setup this Rails application. Initial user accounts can be configured in the db/migrate/20131217194515_create_individuals.rb file.

Operation

The Integrated Biomedical System can be run as a local application with the “rails server” command and a web browser for <http://localhost:3000/> or configured to run as a web application with Apache httpd server. The graphical user interface navigation control panel is a set of eight ovals containing text and image links to interfaces within the application, see top of [Figure 1](#). Users can upload data through the web interface ([Figure S1](#)). A set of command line utilities are included for administrator loading of data ([Table S1](#)).

Results

Interactome

Heart rate monitoring. Heart rate monitoring devices provide heart rate, interbeat interval (IBI), and electrocardiogram (ECG) measurements. Heart rate measurements for multiple devices for an individual are shown in [Figure 1](#). Hidalgo Equivital SEM2 and Zephyr BioHarness were typically worn only during more active periods. Lower Zephyr heart rate values observed on Aug. 29 likely resulted from the contact pads drying out during a period of extended wearing with low activity level. Some data gaps result from the need for device battery recharging (Empatica E3 - daily and Mio Link every 8 to 10 hours). Higher correlations of results are observed for periods of sleeping and light activity. This observation is consistent with previous anecdotal observations of data accuracy and coverage decreases for many wearable sensors during periods of high activity.

Exposome

Sleep monitoring. Multiple devices tested provide top-level estimates of nightly time asleep and number of sleep interruptions. Some devices also attempt to break down the sleep time into sleep phases (light, deep, and rapid eye movement - REM sleep). This data was integrated to enable comparisons of sleep classifications assigned by these devices (investigation of the accuracy of these estimates vs. gold-standard polysomnography was beyond the scope of the present work). Example longitudinal measurements from a single individual collecting data in parallel using Jawbone Up, Basis B1, Fitbit, and ActiSleep are shown in [Figure 2](#). Analytical modules enabling pairwise comparisons of unfiltered nightly time asleep estimates between different devices were developed and integrated into the Integrated Biomedical System. Simple comparisons of daily total time asleep reported across the range of devices revealed a lack of correlation for most device pairs as measured by Pearson r statistics. Likewise, finer-grained estimates of light sleep (provided by Basis and Jawbone) and deep sleep (Jawbone) compared to deep sleep plus REM sleep (Basis) were also poorly correlated. Only the two Actigraph algorithms, Sadeh and Cole-Kripke, which were run on the same raw Actigraph sensor data produced highly correlated results (r of 0.97).

Exposures

Global Position System (GPS) tracking of outside activities available in the Integrated Biomedical System from smartphone or GPS data can provide continuous localization for an individual.

Figure 1. Heart Rate Monitoring. (A) Screen shot of heart rate beats per minute measurements for a volunteer wearing Basis B1 watch, Empatica E3, Zephyr BioHarness, Hildago Equival SEM2, and Mio Link devices. SEM2 values were filtered for minimum quality values of 70 with selection of median value; (B) Zoomed in view of heart rates illustrating measurements at different activity levels; and (C) Bland-Altman plots comparing measurements from the heart rate tracking devices with corresponding Pearson r correlation values.

Figure 2. Sleep Monitoring. (A) Screen shot of daily total sleep measurements for a volunteer for Fitbit Flex, Jawbone Up, Basis B1 watch, and Actisleep. (B) Bland-Altman plots comparing measurements from the sleep tracking devices for this volunteer.

This data enables a range of potentially useful correlations to be determined including correlations with data from nearby EPA or other air quality monitoring station(s) as an initial step toward quantitative tracking of individual exposures. Inferred exposure levels can be estimated from nearby sensors for a wide variety of measured pollutants, particulates⁴², and pollen levels⁴⁵. **Figure 3** illustrates NO₂, PM_{2.5}, carbon monoxide, and ozone exposures for an afternoon walk.

Discussion

Vision

Genome, interactome, and exposome all influence an individual's wellness. The Integrated Biomedical System was developed to demonstrate the ability to begin integrating these heterogeneous data sources in near real-time for individuals. This was accomplished using an architecture that can operate on a stand-alone laptop or desktop personal computer (PC) to provide additional privacy and security and can be connected seamlessly to voluntarily transfer selected data to centralized highly scalable

systems built on the same data architecture that can integrate data from many thousands or even millions of individuals. This approach could provide a path to developing new crowd-sourced models for large-scale prospective/retrospective studies of how individual combinations of genomic and environmental factors correlate with a range of human health and performance traits. Individual monitoring devices, genetic data, blood biochemistries, nutrition, exposures, illnesses, vocal and additional data have been organized and integrated into a unified system. Using the same tools and architectures, additional quantitative lab results and diagnostic data like images and physiological monitoring system data can be added to further increase the research scope of the system. Incorporation of additional natural language processing tools and data architecture modifications can enable text-based metadata collections (e.g. regular symptoms logging from personal health blogs, social interaction details from social media platforms, information from electronic health records) to be included in future versions of the system. Furthermore, these personal datasets can be combined with relevant public datasets

"8.30.2014.walk", 2:11 to 4:05 PM, Aug 30, 2014

Figure 3. Outdoor walk and Integration with EPA AirData. Example visualization of activity data with estimated exposure levels from nearby EPA AirData monitoring site.

and other non-public data to provide new insights into health-associated effects to support detailed N-of-1 and population retrospective analyses.

Genome

As large-scale DNA sequencing costs continue to decrease, sequencing an individual's DNA becomes more affordable and practical. Current costs enable exome sequencing of individuals for less than \$1,000. In a few years, the costs for whole genome sequencing for individuals is projected to be below \$1,000 for very large studies. The quality and completeness of results can be estimated by coverage, but room for improvement is illustrated by the Proton and Illumina exome results correlated with 23andMe SNP profiles. While tools exist to characterize variants (Polyphen2, SIFT, etc.), the potential to correlate variants with protein structure/function, physiology, molecular biomarkers, etc. typically is done manually and within studies with a single focus. Integrating genomic data with interactome and exposome data will help create new opportunities for turning data into new discoveries and knowledge. The Integrated Biomedical System also supports detailed analysis of variant analysis for genes, proteins, pathways, individual SNPs, and other variant types. Future inclusion of raw genomic sequencing data and connections with a variety of genome viewers is straightforward using this extendable data and software architecture. As advances in DNA sequencing technology enable more widespread access to genomic data for individuals, the ability to correlate that data with quantitative interactome and exposome data will become increasingly important. Together, these data can broadly enable efforts to elucidate the interplay between genomic and environmental factors that contribute to complex individual human traits and health.

Interactome

Cognitive performance and health phenotypes can be assessed through a variety of indirect methods including analysis of biomarkers in blood, psychomotor vigilance task (PVT), profile of mood states (POMS), automated neuropsychological assessment metrics (ANAM), speech analysis, facial and eye movement tracking, electroencephalography (EEG), and similar approaches. These assessments and others have been developed and used quantitatively define progressions of important traits/symptoms in individuals experiencing a number of conditions including depression⁴⁶, posttraumatic stress disorder (PTSD), and traumatic brain injury (TBI), as well as environmental stressors including sleep disruption, etc. Data streams produced from these assessments combined with traditional measurements of traits, molecular biomarkers, and clinical data to provide a new platform for gaining insight into the underlying physiology individual health, fitness, and well-being. Retrospective analysis of large-scale collections will provide future biomedical discoveries. Increasing proportions of future biomedical discoveries will be driven by the ability to effectively collect, manage, and interpret massive amounts of heterogeneous data. Enhancements to integrate additional interactome data types and analysis tools

are currently underway and these features will be included in future releases.

Exposome

Asthma and COPD affect 18.7 and 6.8 million individuals in the United States⁴⁷. Environmental exposures can exacerbate these conditions⁴⁸. Asthma can be triggered by particulate matter, ozone, sulfur dioxide, nitrogen oxide, and pollens⁴⁹. Devices, including smart phones, with GPS tracking ability enable the possibility of data integration with environmental monitoring data. Nearby monitoring stations and mobile monitoring devices provide weather and exposure estimates that can be correlated using time stamped GPS positional information. Monitoring stations track a rich variety of environmental exposure data⁴². While the current system provides incomplete coverage, it demonstrates a viable path to incorporation of additional sensor streams (including indoor air quality sensors, UV exposures, etc.) and activity-based estimates of indoor vs. outdoor exposures. It will be possible to provide increasingly complete individualized and integrated quantitative estimates of specific exposures that can be correlated with possible health effects, symptoms, and well-being. Larger and more complete data sets enabled by integrated systems like the one described here, can play a key enabling role for more quantitative genome vs. environment studies in the future.

Conclusions

The Integrated Biomedical System is being developed as an open source platform for individual health, fitness, and in the future wellness promotion. Data visualization, data mining, and new big data approaches will be integrated into the data analysis capabilities that will continue to expand over time. With the goal of creating an open data architecture that supports data exploitation and decision support, this system aims to provide useful information to individuals, medical personnel, researchers, and decision makers. Individuals can run this system on their home computer for use with their own data (and family members). This system will also support longitudinal studies integrating genome, interactome, and exposome heterogeneous data sources. Improving interfaces for user friendliness with valuable feedback and data visualization will be essential for user acceptance, continued use, and progress towards wellness promotion.

Data and software availability

Latest source code: <https://github.com/doricke/IBio>.

Archived source code as at time of publication: <https://doi.org/10.5281/zenodo.1156331>⁵⁰

Software license: GNU General Public License, version 3.0.

The heart rate and sleep tracking data are included in Ricke, Darrell, 2017, "Integrated Biomedical System", doi:[10.7910/DVN/DEEH12](https://doi.org/10.7910/DVN/DEEH12)⁵¹, Harvard Dataverse, V4.

The Equivital SEM data are included in Ricke, Darrell, 2018, "Integrated Biomedical System Equivital SEM", doi:[10.7910/DVN/FD4B6C](https://doi.org/10.7910/DVN/FD4B6C)⁵², Harvard Dataverse, V1.

Competing interests

No competing interests were disclosed.

Grant information

This work is sponsored by the Assistant Secretary of Defense for Research & Engineering under Air Force Contract #FA8721-05-C-002. Opinions, interpretations, recommendations

and conclusions are those of the author and are not necessarily endorsed by the United States Government.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

The authors would like to acknowledge Carl Ricke, Freelance Illustrator at Wanderbots, for graphic artwork.

Supplementary material

Figure S1: Integrated Biomedical System web interface.

[Click here to access the data.](#)

Table S1: Integrated Biomedical System command line utilities for data loading. These extract, transform, and load (ETL) modules provide administrator tools capabilities for bulk loading of data files. Tools are run with the prefix “rails runner lib/utilities/<ETL loader. rb> <parameters>.”

[Click here to access the data.](#)

References

- Hamza TH, Chen H, Hill-Burns EM, *et al.*: **Genome-wide gene-environment study identifies glutamate receptor gene *GRIN2A* as a Parkinson's disease modifier gene via interaction with coffee.** *PLoS Genet.* 2011; 7(8): e1002237.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Wild CP: **Complementing the genome with an “exposome”: the outstanding challenge of environmental exposure measurement in molecular epidemiology.** *Cancer Epidemiol Biomarkers Prev.* 2005; 14(8): 1847–1850.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Gibbs WW: **Medicine gets up close and personal.** *Nature.* 2014; 506(7487): 114–115.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Hood L, Price ND: **Demystifying disease, democratizing health care.** *Sci Transl Med.* 2014; 6(225): 225ed5.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Smarr L: **Quantifying your body: a how-to guide from a systems biology perspective.** *Biotechnol J.* 2012; 7(8): 980–991.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Li-Pook-Than J, Snyder M: **iPOP goes the world: integrated Personalized Omics Profiling and the road toward improved health care.** *Chem Biol.* 2013; 20(5): 660–666.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Chen R, Snyder M: **Systems biology: personalized medicine for the future?** *Curr Opin Pharmacol.* 2012; 12(5): 623–628.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hood L, Auffray C: **Participatory medicine: a driving force for revolutionizing healthcare.** *Genome Med.* 2013; 5(12): 110.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Xu X, Zhu X, Dwek RA, *et al.*: **Structural characterization of the 1918 influenza virus H1N1 neuraminidase.** *J Virol.* 2008; 82(21): 10493–10501.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Martin Sanchez FM, Gray K, Bellazzi R, *et al.*: **Exposome informatics: considerations for the design of future biomedical research information systems.** *J Am Med Inform Assoc.* 2014; 21(3): 386–390.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Doherty ST, Oh P: **A multi-sensor monitoring system of human physiology and daily activities.** *Telemed J E Health.* 2012; 18(3): 185–192.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Nieuwenhuijsen MJ, Donaire-Gonzalez D, Foraster M, *et al.*: **Using personal sensors to assess the exposome and acute health effects.** *Int J Environ Res Public Health.* 2014; 11(8): 7805–7819.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Goldberger AL, Amaral LA, Glass L, *et al.*: **PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals.** *Circulation.* 2000; 101(23): e215–e220.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Montague E, Stanberry L, Higdon R, *et al.*: **MOPED 2.5—an integrated multi-omics resource: multi-omics profiling expression database now includes transcriptomics data.** *OMICS.* 2014; 18(6): 335–343.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Darwin C: **On the Origin of Species.** 1859.
[Publisher Full Text](#)
- David LA, Materna AC, Friedman J, *et al.*: **Host lifestyle affects human microbiota on daily timescales.** *Genome Biol.* 2014; 15(7): R89.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Brito IL, Yilmaz S, Huang K, *et al.*: **Mobile genes in the human microbiome are structured from global to individual scales.** *Nature.* 2016; 535(7612): 435–439.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ormond KE, Wheeler MT, Hudgins L, *et al.*: **Challenges in the clinical application of whole-genome sequencing.** *Lancet.* 375(9727): 1749–1751.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Boeckmann B, Bairoch A, Apweiler R, *et al.*: **The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003.** *Nucleic Acids Res.* 2003; 31(1): 365–370.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Law V, Knox C, Djoumbou Y, *et al.*: **DrugBank 4.0: shedding new light on drug metabolism.** *Nucleic Acids Res.* 2014; 42(Database issue): D1091–D1097.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Just W: **Computational complexity of multiple sequence alignment with SP-score.** *J Comput Biol.* 2001; 8(6): 615–623.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Wang L, Jiang T: **On the complexity of multiple sequence alignment.** *J Comput Biol.* 1994; 1(4): 337–348.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Berman HM, Westbrook J, Feng Z, *et al.*: **The Protein Data Bank.** *Nucleic Acids Res.* 2000; 28(1): 235–242.
[PubMed Abstract](#) | [Free Full Text](#)
- Prosite database.**
[Reference Source](#)
- Fitch WM, Langley CH: **Protein evolution and the molecular clock.** *Fed Proc.* 1976; 35(10): 2092–2097.
[PubMed Abstract](#)

26. Ricke D, Shcherbina A, Chiu N, *et al.*: **Sherlock's Toolkit: A forensic DNA analysis system**. Technologies for Homeland Security (HST), IEEE International Symposium on. 2015.
[Publisher Full Text](#)
27. Ricke DO: **BioTools**. Bioinformatics programs.
[Reference Source](#)
28. Gribskov M, McLachlan AD, Eisenberg D: **Profile analysis: detection of distantly related proteins**. *Proc Natl Acad Sci U S A*. 1987; **84**(13): 4355–4358.
[PubMed Abstract](#) | [Free Full Text](#)
29. Shcherbina A, Ricke DO, Schwoebel E, *et al.*: **KinLinks: Software Toolkit for Kinship Analysis and Pedigree Generation from HTS Datasets**. Technologies for Homeland Security (HST), IEEE International Symposium on. 2016.
[Publisher Full Text](#)
30. **Jmol: an open-source Java viewer for chemical structures in 3D**.
[Reference Source](#)
31. Sigrist CD, Cerutti L, Hulo N, *et al.*: **PROSITE: a documented database using patterns and profiles as motif descriptors**. *Brief Bioinform*. 2002; **3**(3): 265–274.
[PubMed Abstract](#) | [Publisher Full Text](#)
32. Whittle JR, Zhang R, Khurana S, *et al.*: **Broadly neutralizing human antibody that recognizes the receptor-binding pocket of influenza virus hemagglutinin**. *Proc Natl Acad Sci U S A*. 2011; **108**(34): 14216–14221.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
33. UniProt Consortium: **Activities at the Universal Protein Resource (UniProt)**. *Nucleic Acids Res*. 2014; **42**(Database issue): D191–D198.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
34. Ricke DO: **Analysis of Sequence and Molecular Evolution Information in Two Model Systems**. Mayo Graduate School. 1995.
35. Bottema CD, Ketterling RP, Vielhaber E, *et al.*: **The pattern of spontaneous germline mutation: relative rates of mutation at or near CpG dinucleotides in the factor IX gene**. *Hum Genet*. 1993; **91**(5): 496–503.
[PubMed Abstract](#) | [Publisher Full Text](#)
36. Koeberl DD, Bottema CD, Buerstedde JM, *et al.*: **Functionally important regions of the factor IX gene have a low rate of polymorphism and a high rate of mutation in the dinucleotide CpG**. *Am J Hum Genet*. 1989; **45**(3): 448–457.
[PubMed Abstract](#) | [Free Full Text](#)
37. Povolotskaya IS, Kondrashov FA: **Sequence space and the ongoing expansion of the protein universe**. *Nature*. 2010; **465**(7300): 922–926.
[PubMed Abstract](#) | [Publisher Full Text](#)
38. Ashley EA: **The precision medicine initiative: a new national effort**. *JAMA*. 2015; **313**(21): 2119–20.
[PubMed Abstract](#) | [Publisher Full Text](#)
39. Ricke DO: **Divergence Model of Protein Evolution**. 2016.
[Publisher Full Text](#)
40. **AMD Opteron 6282 Specint 1250**. 2011.
[Reference Source](#)
41. **Intel Xeon 2698 v3 Specint 1250**. 2006.
[Reference Source](#)
42. Sommer SS, Ketterling RP: **The factor IX gene as a model for analysis of human germline mutations: an update**. *Hum Mol Genet*. 1996; **5**(Supplement 1): 1505–1514.
[PubMed Abstract](#) | [Publisher Full Text](#)
43. **Illumina ForenSeq DNA Signature Prep Kit**. 2016.
[Reference Source](#)
44. Samani NJ, Tomaszewski M, Schunkert H: **The personal genome—the future of personalised medicine?** *Lancet*. 2010; **375**(9725): 1497–1498.
[PubMed Abstract](#) | [Publisher Full Text](#)
45. Gribskov M, McLachlan AD, Eisenberg D: **Profile analysis: detection of distantly related proteins**. *Proc Natl Acad Sci U S A*. 1987; **84**(13): 4355–4358.
[PubMed Abstract](#) | [Free Full Text](#)
46. Williamson JR, Quatieri TF, Helfer BS, *et al.*: **Vocal biomarkers of depression based on motor incoordination**. In: *Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge*. ACM: Barcelona, Spain. 2013; 41–48.
[Publisher Full Text](#)
47. Sahini L, Tempczyk-Russell A, Agarwal R: **Large-scale sequence analysis of hemagglutinin of influenza A virus identifies conserved regions suitable for targeting an anti-viral response**. *PLoS One*. 2010; **5**(2): e9268.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
48. Ko FW, Hui DS: **Air pollution and chronic obstructive pulmonary disease**. *Respirology*. 2012; **17**(3): 395–401.
[PubMed Abstract](#) | [Publisher Full Text](#)
49. Guarneri M, Balmes JR: **Outdoor air pollution and asthma**. *Lancet*. 2014; **383**(9928): 1581–1592.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
50. Ricke D: **doricke/IBio: Integrated Biomedical System (Version 1.0.1)**. *Zenodo*. 2018.
[Data Source](#)
51. Ricke D: **Integrated Biomedical System**. *Harvard Dataverse*, V5. 2017.
[Data Source](#)
52. Ricke D: **Integrated Biomedical System Equivital SEM**. *Harvard Dataverse*, V1. 2018.
[Data Source](#)

Open Peer Review

Current Peer Review Status:

Version 1

Reviewer Report 06 April 2018

<https://doi.org/10.5256/f1000research.14774.r32216>

© 2018 Kuchinke W. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Wolfgang Kuchinke

Coordination Centre for Clinical Trials, Heinrich Heine University Düsseldorf (HHU), Düsseldorf, Germany

Integrated Biomedical System

This article addresses a very important topic, the integration of different kinds of data from different domains for joint analysis. This is indeed the future of research to jointly analyze genomic, clinical, life style and environmental data. The authors claim to have developed an "Integrated Biomedical System", a platform that collects and stores in a single place different data sources, like genome data (SNPs, NCBI), proteomic data (PDB, SwissProt), physiologic data (heart rate, skin conductance, respiratory rate derived from sensors), life style data (nutrition, prescriptions, other medications, sleep pattern) and air pollution data from EPA and used it for some kind of study. This is a considerable achievement, but it is presented in an unclear, incomplete and insufficient way.

To begin, what is this manuscript? A description of a new software tool, a study of the integration of sensor data, a pilot of a software tool? The purpose of the manuscript should be clearly stated; it determines how the results have to be presented.

An important aspect of such a study is how data is measured, treated and processed. There are different methods and algorithms used to analyze data from sensors used for monitoring. For example, sensor signals can be contaminated with noises or interferences. In addition, responses to these signals could be different for different persons. In this context, sensor data need calibration. All these issues are not discussed. In fact, the integration of data from diverse sources results in data with different data formats, data models, metadata schema, etc. and this must be addressed.

Efforts of other groups working on the integration of sensor and physiological data should be mentioned. For example, Arturo Arino from the University of Navarra combines in the PAIRQURS project physiological data with location data and air pollution data measured with sensors

attached to bicycles [1]. These sensors record the levels of selected atmospheric pollutants, like CO, NOx, Ozone, and airborne particles, together with auxiliary data (Temp, HR) and GPS coordinates and transmit processed packets via GPRS to a central database.

In detail:

Methods

The Methods section is very weak. What is the *Experimental Procedure*?

Protocol design and approvals are mentioned; but is it a medical study? What is the study design? What is the comparator? It is stated that the study was planned for 20, later 40 participants. But how many volunteers actually participated, what gender, age, number of drop-outs, ...?

Data *privacy protection* is a hot topic for medical data and GPS data. It should have been considered in the informed consent, because with GPS data identifiability of persons becomes a problem. Privacy protection should have been discussed.

"Volunteers have full choice of all elements of the research project ..." What does this mean? Can participants determine the study protocol?

There is sometimes redundancy in the text:

For example, "COUHES approval", "Entering events, meals, prescriptions..." is mentioned 2 times

It seems that some terms are used arbitrarily. For example, the "*Genome* part" contains also Proteome data. "Interactome" means the whole set of molecular interactions in a cell, like the human protein-protein interactome. But in the manuscript Interactome contains even heart rate, ECG, respiratory rate, etc. The Exposome encompasses the totality of human environmental exposures. But in the manuscript it contains measurements of activity and sleep and nutrition.

The Integrated Biological System is obviously a software tool, including several databases. But no exact information about the structure and architecture of the system is given. How many tools / modules, what interfaces, how many databases, what data models, what standards, etc. Global unique identifiers are mentioned. How are these identifiers integrated in the system? A GUID is a number that the program generates to create a unique identity for an entity; why is this used for data exchange between sites (and between which sites).

Implementation. This section contains partly redundant information with Integrated Biological System, describing that Ruby on Rails was used, some of the databases, etc. Better a list should be provided that describes all components of the application setup and their connections.

Operation

Concerns the use of the Integrated Biological System. "Users can upload data through the web interface". This is very basic information, without any details. But what data, what format, how to use the interface, ...?

Measurement procedures are missing. What are the Analytical Methods? Statistical Analyses? Some information is included in Fig. 1 (like, Bland-Altman plot is mentioned, but why use this plot?)

Results

The Results section has the same subtitles as the Methods section. But here the results of the pilot study should be presented in a standardized and comprehensive way. First, what is the primary result of the pilot study? The detailed numeric outcome could be summarized in a table.

Many problems are mentioned, like devices worn only during active periods, contact pads drying out, battery recharging, etc. What was the effect on the data? This should be mentioned. In addition, an interpretation of Fig.1 is missing. The reason for comparing such a large number of different devices is missing. Were different devices worn by the same person, were multiple measurements done?

Several rather vague statements, for example:

"multiple devices were tested .." how many?

"Some devices also ..." how many?

"This data was integrated to enable ..." how?

"... provided by a subset of volunteers." How many?

"Example longitudinal measurements from a single individual .." how was this person selected?

"Analytical modules were developed ..." belongs to Methods

"GPS tracking of outside activities ..." how could outside activities be differentiated from inside activities?

"...data from nearby EPA or other air quality monitoring stations." More information is necessary. How were the monitoring stations selected, how far away are these stations from the participant, what kind of data is provided by the stations, what kind of calibration was used, ...?

Discussion

Here the results should be discussed in a critical way. This is completely missing in the manuscript. The vision should be put to the final end of the manuscript.

It is stated that individual monitoring devices, genetic data, blood biochemistries, nutrition, exposures, ... were integrated into a unified system. But the integration was not complete, because the genetic data seems to be missing in the results section. Important is to discuss how good this integration was, and all results should be evaluated critically.

Reference

<https://eudat.eu/news/interview-with-arturo-ari%C3%B1o-from-pairqurs>

Is the work clearly and accurately presented and does it cite the current literature?

No

Is the study design appropriate and is the work technically sound?

No

Are sufficient details of methods and analysis provided to allow replication by others?

No

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

No

Are the conclusions drawn adequately supported by the results?

No

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Clinical research and IT infrastructures

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Reviewer Report 26 February 2018

<https://doi.org/10.5256/f1000research.14774.r30676>

© 2018 Lopez-Campos G et al. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Guillermo H. Lopez-Campos

Wellcome-Wolfson Institute for Experimental Medicine, Queen's University of Belfast, Belfast, UK

Philip Kiossoglou

Health and Bioemdicinal Informatics Centre, The University of Melbourne, Parkville, VIC, Australia

This article refers to the development of the “Integrated Biomedical System” (IBio) a system developed as a freely available Ruby on Rails application. The proposed system is multiplatform and capable of storing different data sources (genome, “interactome” and exposome) that can be used both at the individual level or to collect data from large populations. For this manuscript, the authors described the use of data gathered through different wearable devices to monitor some physiological parameters, such as heart rate or sleep patterns, and GPS data gathered from the MyTrack app and how they were uploaded into IBio. The development of systems similar to the one described in this manuscript is an interesting step for the future integration and analysis of multiple data sources.

The aim of the paper is to develop an open source platform to integrate different data sources “providing a unifying model to promote more open data sharing and analysis”. Unfortunately the manuscript fails to present evidence of having successfully achieved that aim or how this system would solve some of the problems mentioned in the introduction associated with other projects such as using “a relatively narrow set of measurements, or on custom data storage and analysis architectures that do not provide a scalable foundation for larger scale integration across studies to enable meta-analysis of data from multiple studies” (Ricke DO et al. F1000Research 2018). The manuscript does not describe the suggested unifying model and it is limited (because it did not mention the use of any standards) to a set of “ad hoc” implemented scripts/modules to a limited

set of devices and measurements and does not show any evidence of having used any "genomic data". It neither provides neither descriptions nor evidence of how IBio promotes more open data sharing and analysis. Data integration is limited to store the data in the same platform but keeping the three different data types separated from each other. The results section is mostly focused in comparisons between the different devices rather than in the system itself. Because of these reasons and despite working in an extremely interesting area this manuscript cannot be approved unless it undergoes major and extensive revisions in its design and its contents.

Major revisions

1.

The article should provide a much better and extensive analysis of the existing literature and approaches and how these data are integrated or stored in the same repository. This analysis would facilitate identifying the existing alternatives, their deficiencies and therefore would put in context the contribution of the proposed solution. An example of relevant existing contributions would be the work developed by the National Center of Excellence for Mobile Sensor Data-to-Knowledge (MD2K) (1) but also other approaches such as those described in the references 2 to 7. In addition, other interesting approaches are the initiative known as "WikiLife" and the UK Biobank (<http://www.ukbiobank.ac.uk>). Wikilife aimed to manage and integrate different data sources and has left some software (<https://github.com/wikilife-org>), which could be compared with the system described in this manuscript. The UK Biobank also contains information similar to the one stored in IBio (It does not contain GPS data but does contain some other sources of geographical information that might be used to input exposure data) and its structure might be useful to compare both systems.

From a technical and technological perspective for data integration and data sharing the manuscript did not mention any standardization initiatives or attempts. The authors should reflect upon this issue. A relevant example in this direction is the work developed by Open mHealth (<http://www.openmhealth.org/>) (8,9). These are all aspects that should be included in the paper and considered in the discussion or interpretation of the results as elements of comparison.

2.

The use of the term "interactome" in this manuscript is very confusing. It is widely accepted that the "interactome" is the set of interactions among proteins (10) whereas the authors have employed it in this manuscript in the context of physiological parameters or phenotypes, making it difficult to relate to the actual contents and data. Therefore, it should be replaced across the whole text with some other term that better describes these elements.

3.

The methods section provides some information about some of the data contained in the system, and how some volunteers gathered some of the data. Despite the manuscript mentioning the use of genome data by iBio, the methods section only reflects that they store information from other sources. The authors did not mention how the information from the different resources was integrated and related. This is critically relevant information that should be included in the manuscript.

The manuscript did not provide any criteria for the selection or the inclusion of the different devices & data used to build the system.

The "interactome" section describes elements designed to interact with the Basis Science website for the collection of these data however the company stopped offering their services on 31st

December 2016 thus those modules are useless and therefore they can be removed from the text.

4.

The exposome subheading in the methods section refers mostly to sleep data gathered using some wearable devices (the some previously described in the "interactome" section). These data are not exposure data but physiological data. Other proper exposure data such as diet and prescriptions are poorly described and not referred at later stages with the exception of GPS data. Information about how these data are managed and stored in the system is relevant and should be provided. V.g. Were they entered just as free text?

5.

The results are not clearly related to the aims of the manuscript and should be revisited to better present the actual results of their research. A large part of the results presented actually relate to comparisons between the devices used by the participants rather than about the platform itself. The results section seems to focus more in the research question "Do different devices produce the same results?" than in the integration aspects which allegedly are driving the research and the manuscript. Methods section should contain information about the methodology used in these comparisons.

As the manuscript is describing a system based in a database the structure and contents of this database should have been presented in the results section. Results should have focused as well the interfaces developed for the system, user experience or the comparison with other existing platforms.

One of the main aspects of the manuscript is data integration, however with the results presented in the manuscript it is unclear the benefit of having the data in this platform or having three different platforms storing the same data

6.

The results section strikingly lacks results about the integration of any genomic data despite frequent mention of these data in the manuscript. For testing purposes the authors should at least enter some data (that could be collected from online resources) or could be simulated.

7.

Surprisingly, the authors distinguish between exposome and exposure data in the results section, what is the reason for that? By definition the exposome is the whole set of exposures of an individual. A detailed reading of this section shows that it just reflects about the possibility of including GPS data into the system that potentially could be linked with other sources such as the EPA. This is an interesting idea but sadly the manuscript lacks details about how this integration is performed and it should be included in the methods section.

8.

The discussion section is subdivided in four different elements, (vision, genome, interactome, exposome). Overall, the discussion very vaguely discusses the results presented in the paper, instead focusing on potential future applications and listing potential data sources. The results section should be revisited to actually focus on the discussion of the results.

The discussion starts with a vision that talks about genomic data, however as previously

mentioned there is no evidence in the manuscript that any data of this kind has been uploaded into the system but these data are not mentioned anywhere in the manuscript other than in the methods section.

This paragraph also mentions that the system demonstrates the ability to integrate data in near real-time. This is not evident in the manuscript. As the system requires manual input of the data by the user it is unclear how this can be considered “near real-time”. These statements should be therefore corrected to better reflect the reality of the system as presented in this manuscript.

The “Genome” subheading in the results section must be removed or heavily amended. It provides a vague description of what can be done with genomic data and lacks evidence to support any other of the claims made. It is important to mention that the manuscript does not provide any references about how any genomic data can be uploaded into the system and therefore does not actually discuss any results. It also contains some vague references to some tools (that require a proper citation) that are not included in the system and there is an absolute lack of evidence that the system would be able to support any detailed analysis of any molecular data (genetic, protein, pathways...). In the current text, it is unclear what difference this system would make in terms of the analysis of the different variants in terms of automating these analyses or broadening their focus.

9.

The conclusions are vague and should be edited to better reflect the status of the system and the actual conclusions from the work presented. For example, according to the manuscript, it is unclear how without adequate integration this system would be able to provide useful information in medical environments (for medical personnel) or decision makers without analysing security or data sharing issues. Other conclusions are based in future improvements done in the system which are not available yet or haven't been described and in an integration that at the moment does not go any further than storing data together in the same platform without actually integrating them.

The authors themselves acknowledge that the system requires improvements in the user interface, aspects that should have been addressed much earlier in the manuscript and that are key for the success of these kind of applications.

10.

Surprisingly, a large number of the citations are incorrect and do not refer or match with the contents in the manuscript. (e.g. references 8, 15, 18,21,22,24,25,26,27....).

Minor corrections

1.

As a minor element in this aspect notice that reference 10 should read as Martin-Sanchez et al rather than Sanchez et al.

2.

In the methods section authors said that in some cases the volunteers underwent a microbiome analysis but later on there are no mentions of this data nor how they can be incorporated therefore it should be removed from the manuscript.

3.

In Figures 1A and S1 there are elements that are not mentioned at all in the manuscript or being part of the system (For example miRNAs, Ancestry, Pathogens).

4.

Apparently, there is some sort of colour code used in the interface, it would be interesting to provide a description of that...

5.

As previously indicated the “interactome” subheading in the discussion should be modified to better describe the results rather than vaguely talk about potential benefits derived from data integration.

References

1. Kumar S, Abowd G, Abraham WT, al'Absi M, et al.: Center of Excellence for Mobile Sensor Data-to-Knowledge (MD2K).*IEEE Pervasive Comput.* **16** (2): 18-22 [PubMed Abstract](#) | [Publisher Full Text](#)
2. Ng K, Kakkanatt C, Benigno M, Thompson C, et al.: Curating and Integrating Data from Multiple Sources to Support Healthcare Analytics.*Stud Health Technol Inform.* 2015; **216**: 1056 [PubMed Abstract](#)
3. Mezghani E, Exposito E, Drira K, Da Silveira M, et al.: A Semantic Big Data Platform for Integrating Heterogeneous Wearable Data in Healthcare.*J Med Syst.* 2015; **39** (12): 185 [PubMed Abstract](#) | [Publisher Full Text](#)
4. de Arriba-Pérez F, Caeiro-Rodríguez M, Santos-Gago JM: Collection and Processing of Data from Wrist Wearable Devices in Heterogeneous and Multiple-User Scenarios.*Sensors (Basel).* 2016; **16** (9). [PubMed Abstract](#) | [Publisher Full Text](#)
5. Blaauw FJ, Schenk HM, Jeronimus BF, van der Krieke L, et al.: Let's get Physiqua - An intuitive and generic method to combine sensor technology with ecological momentary assessments.*J Biomed Inform.* 2016; **63**: 141-149 [PubMed Abstract](#) | [Publisher Full Text](#)
6. Bai J, Shen L, Sun H, Shen B: Physiological Informatics: Collection and Analyses of Data from Wearable Sensors and Smartphone for Healthcare.*Adv Exp Med Biol.* 2017; **1028**: 17-37 [PubMed Abstract](#) | [Publisher Full Text](#)
7. Kumari P, Lopez-Benitez M, Gyu Myoung Lee, Tae-Seong Kim, et al.: Wearable Internet of Things - from human activity tracking to clinical integration.*Conf Proc IEEE Eng Med Biol Soc.* **2017**: 2361-2364 [PubMed Abstract](#) | [Publisher Full Text](#)
8. Chen C, Haddad D, Selsky J, Hoffman JE, et al.: Making sense of mobile health data: an open architecture to improve individual- and population-level health.*J Med Internet Res.* 2012; **14** (4): e112 [PubMed Abstract](#) | [Publisher Full Text](#)
9. Estrin D, Sim I: Health care delivery. Open mHealth architecture: an engine for health care innovation.*Science.* 2010; **330** (6005): 759-60 [PubMed Abstract](#) | [Publisher Full Text](#)
10. De Las Rivas J, de Luis A: Interactome data and databases: different types of protein interaction.*Comp Funct Genomics.* 2004; **5** (2): 173-8 [PubMed Abstract](#) | [Publisher Full Text](#)

Is the work clearly and accurately presented and does it cite the current literature?

No

Is the study design appropriate and is the work technically sound?

No

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Not applicable

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

No

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Biomedical informatics, exposome informatics, translational bioinformatics

We confirm that we have read this submission and believe that we have an appropriate level of expertise to state that we do not consider it to be of an acceptable scientific standard, for reasons outlined above.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research