Bioinformatics Advance Access published September 1, 2014

BIOINFORMATICS APPLICATIONS NOTE 2014, pages 1–3
doi:10.1093/bioinformatics/btu554

Databases and ontologies

Advance Access publication August 22, 2014

ProtocolNavigator: emulation-based software for the design, documentation and reproduction biological experiments

Imtiaz A. Khan[1,2,*], Adam Fraser[2], Mark-Anthony Bray[2], Paul J. Smith[1], Nick S. White[3], Anne E. Carpenter[2] and Rachel J. Errington[1]

[1]School of Medicine, Cardiff University, Cardiff, UK, [2]Imaging Platform, Broad Institute of MIT and Harvard, Cambridge, MA, USA and [3]School of Optometry and Vision Sciences, Cardiff University, Cardiff, UK

Associate Editor: Jonathan Wren

ABSTRACT

Motivation: Experimental reproducibility is fundamental to the progress of science. Irreproducible research decreases the efficiency of basic biological research and drug discovery and impedes experimental data reuse. A major contributing factor to irreproducibility is difficulty in interpreting complex experimental methodologies and designs from written text and in assessing variations among different experiments. Current bioinformatics initiatives either are focused on computational research reproducibility (i.e. data analysis) or laboratory information management systems. Here, we present a software tool, ProtocolNavigator, which addresses the largely overlooked challenges of interpretation and assessment. It provides a biologist-friendly open-source emulation-based tool for designing, documenting and reproducing biological experiments.

Availability and implementation: ProtocolNavigator was implemented in Python 2.7, using the *wx* module to build the graphical user interface. It is a platform-independent software and freely available from http://protocolnavigator.org/index.html under the GPL v2 license.

Contact: wpciak@cf.ac.uk

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on May 23, 2014; revised on July 29, 2014; accepted on August 11, 2014

1 BACKGROUND

Reproducibility is a fundamental tenet for all scientific endeavour that ensures the credibility of findings. It is particularly important for data serving as the basis for the development of therapeutics, an extremely costly process. Recent studies show an alarming increase of irreproducible biological research (Begley and Ellis, 2012) and decrease of biological data reuse (Editorial, 2011).

Prominent reproducibility initiatives in bioinformatics are focused either on Taverna-like (Wolstencroft *et al.*, 2013) 'data analysis' centric computational research or laboratory information management systems like 'data curation-management' infrastructures (Rocca-Serra *et al.*, 2010). Much less emphasis has been given to facilitating the capture, interpretation and identification of methodology variation of biological research (Editorial, 2013). For 'long tail' biologists (Wallis *et al.*, 2013), where the majority of biological experiments are undertaken, the practice variation (how individual researchers have executed the method) among experimentalists remains idiosyncratic and changes frequently. In the absence of a well-structured methodology or standard operating procedure, the design of an experiment is crucial for interpretation and identification of variability. Therefore, illustrating the design in an understandable and flexible manner is critical when addressing the challenges of reproducibility (Millard *et al.*, 2011). However, the experimental design is an abstract representation of the strategies undertaken, and at present, is conveyed by written narratives or interpersonal conversations. Both these approaches are unreliable, as they depend on a shared language, overlapping experiences and alignment of mental models (Eriksson and Webster, 2008).

2 RESULTS AND DISCUSSION

Addressing these challenges, here we introduce ProtocolNavigator. Centred on cell biology research, this software provides an interactive environment where experimentalists can emulate their laboratory practice on a virtual laboratory bench. As a result of 'acting out' the experiment, the design of the experiment is automatically depicted onto a canvas as a map with action icons and tracks for navigation. It is thus more readily adopted than software designed to simply capture metadata, which requires filling in forms or building workflows, and yet provides no immediate understanding about the experimental design and sample handling processes.

ProtocolNavigator consists of three panels with linked functionality and display. Using the **Inventory panel**, the user creates an 'inventory' of instances with detailed descriptions of items such as instrumentation, materials and reagents. Importantly, this inventory can be reused, adapted and shared. Using the **Bench panel** (similar to a laboratory workbench), the previously created instances from the inventory are applied to different experimental samples at certain time points. The time-integrated action-based documentation approach in ProtocolNavigator is unique in enabling researchers to capture their real-life laboratory practice (e.g. temporal variation of activity). The **Map panel** automatically depicts the practice map or design of the experiment, with spatiotemporally linked set of branches and activity icons for retrieving both action details and experiment-derived data. The linking of laboratory practice with actual experimental

*To whom correspondence should be addressed.

data on a jargon-independent map inherently provides a good foundation for different disciplines to communicate and identify experimental design and the underlying practice variation—an essential requirement for reproducibility.

Importantly, the fully navigable map can be shared with colleagues and, therefore, introduces a unique capacity for *in silico* collaborative and coordinated experimental design optimization and planning, potentially reducing experiment iterations and associated costs. The map can be converted and printed into a time-stamped sequential description of steps suitable for carrying out the procedure at the physical bench or for publication. The data file underpinning the map can be easily parsed and reformatted.

3 CASE STUDY

3.1 Measurement of nanoparticle redistribution in a proliferating tumour population - establishing the principle of a 'conserved signal'

The researchers' goal is to establish the appropriate experimental designs to determine how a biophotonics signal from the nanoparticles is diluted over time as a tumour population proliferates. Two alternate protocols were designed using ProtocolNavigator and discussed among the interdisciplinary research team.

In design terms, the first approach (Fig. 1A) achieves a time course for the tumour cultures achieved by seeding the cells at staggered time points (on sequential days), and labelling each population with nanoparticles 24 h post seeding. Thus, all the samples are delivered on the same day for final measurement using a flow cytometer. The second approach (Fig. 1B) uses a continuous sampling approach, where cells are seeded identically in different vessels at the single time point and the cultures are labelled and processed for measurement using flow cytometry on five subsequent days. In both cases, the cell culture is sampled after 2, 3, 4 and 5 days of growth and, in both cases, the biologist is able to adequately monitor the growth parameters of the cultures as well as the attenuation of nanoparticle signals (fluorescence per cell). The staggered approach is classically used when the perturbing agent to be added is a small molecule or 'drug', where uptake by the cell is assumed to be invariant each time. It might also be preferred if it is important for data acquisition to occur on the same day for convenience or to ensure control of instrument performance.

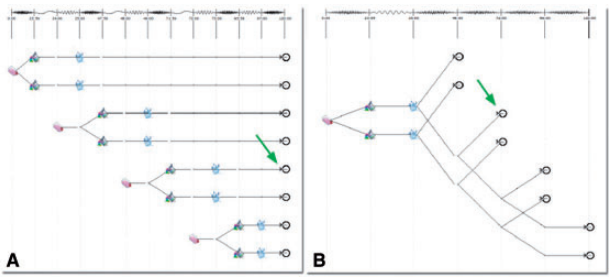

Fig. 1. ProtocolNavigator screenshot showing two alternate experimental designs for the labelling and tracking of tumour systems. *Staggered* seeding (**A**) versus *Continuous* sampling (**B**). (A) Cells were seeded (stock culture flask icon) and then labelled (paint icon) with fluorescently labelled nanoparticles, including a corresponding unlabelled control. The biology and cell labelling are both established in a staggered manner. Cell seeding and labelling was done in a paired fashion at 0–23.59 h; 24–47.59 h; 48–71.59 h; 72–95.59 h, followed by wash (water droplet icon) to remove the remnant labels from the medium. Flow cytometer data acquisition (laser icon) was performed for all eight vessels (tubes) on the 5th day, i.e. at 120 h. (B) Here, cells were seeded in all eight vessels at time 0 h and then labelled with nanoparticles (four vessels; 23.59 h later) again with corresponding unlabelled control (four vessels). Thereafter, data acquisition was performed on vessels (one from fluorescently labelled and one from control) 48, 72, 96 and 120 h post seeding. For both scenarios 'A' and 'B', eight flow cytometer FCS files were acquired of cells exposed to nanoparticles and control for varying numbers of days. Experiments are available in Supplementary Information for download and navigation through ProtocolNavigator

However, from a nanoparticle perspective there is a fundamental flaw in the staggered design: the uptake of 'particulates' by cell is innately variable (Summers *et al*., 2011), and as a result, the effective cellular labelling is not uniform. This presents unpredictability of the initial nanoparticle uptake per cell, and thus the starting signal, leading to a misinterpretation of the consequential signal processing and implementation of the 'conserved signal' principle. Thus, a continuous sampling approach (Fig. 1B) was chosen, so that the computational researchers could apply this signal conservation principle. The experiment was conducted following the designed protocol, the resulting flow cytometry data (FCS files) was linked to the map, and the map and data were shared with the computational members of the team for successful analysis.

This experience confirmed that for the computational researchers to easily interpret and use these experiment-derived data, visual perception and understanding of the experimental design and provenance information was critical. For example, understanding the crucial differences between the arrow-marked datasets (derived after 3 days of nanoparticle labelling in both experimental designs) was more apparent with visual representation compared with only a narrative text-based description of the experimental metadata.

4 CONCLUSION

Although ProtocolNavigator has been primarily developed for basic cell-based research, new instruments or materials can be added as per prospective users' demand. Importantly, the emulation-based documentation and automatic design depiction concept introduced here is applicable across many research environments where practice, design and interdisciplinary communication are serious concerns (e.g. clinical trials or animal studies). By linking practice variation with experimental data, ProtocolNavigator not only introduces an intuitive mechanism for identifying key factors for reproducibility but also the foundation to convey best practices in quantitative terms.

Funding: This work was supported by European Union Marie Curie International Outgoing Fellowship (FP7-PEOPLE-2009-IOF-254046) and Union for International Cancer Control Fellowship (YY2/08/011) to I.A.K. This work was further supported by the US National Institute of Health (R01 grant GM089652 to A.E.C), Human Frontiers in Science Program

(A.E.C) and UK Engineering and Physical Sciences Research Council grant (EP/G037841/1 to R.J.E).

Conflict of interest: none declared.

REFERENCES

Begley,C.G. and Ellis,L.M. (2012) Drug development: raise standards for preclinical cancer research. *Nature*, **483**, 531–533.

Editorial. (2011) Challenges and opportunities. *Science*, **331**, 692–693.

Editorial. (2013) Reducing our irreproducibility. *Nature*, **496**, 398.

Eriksson,L. and Webster,A. (2008) Standardising the unknown:practicable pluripotency as doable futures. *Sci. Cult.*, **17**, 57–69.

Millard,B.L. *et al.* (2011) Adaptive informatics for multifactorial and high-content biological data. *Nat. Methods*, **8**, 487–493.

Rocca-Serra,P. *et al.* (2010) ISA software suite: supporting standards-compliant experimental annotation and enabling curation at the community level. *Bioinformatics*, **26**, 2354–2356.

Summers,H.D. *et al.* (2011) Statistical analysis of nanoparticle dosing in a dynamic cellular system. *Nat. Nanotechnol.*, **6**, 170–174.

Wallis,J.C. *et al.* (2013) If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. *PLoS One*, **8**, e67332.

Wolstencroft,K. *et al.* (2013) The Taverna workflow suite: designing and executing workflows of Web Services on the desktop, web or in the cloud. *Nucleic Acids Res.*, **41**, 557–561.