CellPress

An open data ecosystem for cell migration research

Paola Masuzzo[1,2], Lennart Martens[1,2], and The 2014 Cell Migration Workshop Participants[3]

[1] Department of Medical Protein Research, VIB, A. Baertsoenkaai 3, 9000 Ghent, Belgium
[2] Department of Biochemistry, Ghent University, A. Baertsoenkaai 3, 9000 Ghent, Belgium
[3] See Contributing Authors section at the end of this paper

Cell migration research has recently become both a high content and a high throughput field thanks to technological, computational, and methodological advances. Simultaneously, however, urgent bioinformatics needs regarding data management, standardization, and dissemination have emerged. To address these concerns, we propose to establish an open data ecosystem for cell migration research.

Where is the cell migration field migrating to?

Cell migration is crucial in biological processes such as morphogenesis, immune surveillance, wound healing, and cancer metastasis [1]. Diverse biological models have been developed to reflect the range of molecular and physiological events involved in cell migration (see Figure S1 in the supplementary material online). Furthermore, technology has been an important driver for innovation in cell migration research. For example, the evolution of light microscopy from bright field to confocal, two photon, light sheet, and superresolution fluorescence microscopy has enabled the development of complex experimental systems, progressing from 2D cell migration assays to 2.5D and 3D (see Glossary) approaches [2] (see Table S1 in the supplementary material online).

While analyses on 2D substrates have led to essential insight into the cellular motility machinery, 3D environments are essential for understanding their physiological context, and have recently provided novel knowledge regarding invasive behaviour [3]. Although these *in vitro* assays are clearly valuable, deeper insight into cell migration can only be obtained through *in vivo* approaches. Such assays have been enabled through live cell microscopy to visualize moving cells in their native surroundings, revealing previously unsuspected feedback mechanisms [4]. Moreover, results in high content and high throughput microscopy have established the importance of quantitative analysis for systems biology and drug discovery [5].

A key remaining challenge is to understand how the function and signalling of organelles is coordinated and integrated within cells and tissues. Cell migration is the product of complex processes operating at different scales, and could be investigated using a systems microscopy approach [6], combining image analysis at different resolutions with data mining, multivariate statistics, and modelling.

These advances in techniques and biological models have been supported by dedicated efforts in bioinformatics and computational biology (see Table S1 in the supplementary material online). Algorithms and tools have been developed for tracking cells using time lapse images [7], and for processing and visualizing large sets of complex image data (http://jcb-dataviewer.rupress.org). The computational approaches in the field extend to *in silico* modelling of cell migration and invasion, especially in tumour development and progression [8]. Advances in the field have thus been built on a combination of novel analytical approaches, dedicated software tools and algorithms, and predictive theoretical models.

Taking on the challenges: an open data ecosystem for cell migration

Even though the cell migration field has embraced computational models as a means to integrate and interpret experiments, a key missing element is the global iterative connection between experimental data and computational approaches. This connection requires an open and free data ecosystem, where standardized and documented results of cell migration research can be shared and consulted within a central location, as exemplified in Figure 1. Building such an ecosystem will require several interdigitated and essential developments. A public, centralized repository constitutes the major component; however, it is only viable if supported by standard formats for the stored data and metadata. Furthermore, each data set in the repository should conform to minimum reporting requirements that ensure consistent annotation (see Table S1 in the supplementary material online). The following sections describe each of these aspects in more detail.

Data and metadata standardization

Minimum reporting requirements

To be reusable, an experimental data set needs accompanying metadata, describing both biological and

Glossary

2D: two-dimensional.
2.5D: two-and-a-half-dimensional.
3D: three-dimensional.
CMC: cell migration consortium.
CMG: cell migration gateway.
CV: controlled vocabulary.
OME: open microscopy environment.

Corresponding author: Martens, L. (lennart.martens@vib-ugent.be).
Keywords: cell migration; bioinformatics; standardization; meta-analysis data ecosystem.

0962-8924/
 http://dx.doi.org/10.1016/j.tcb.2014.11.005

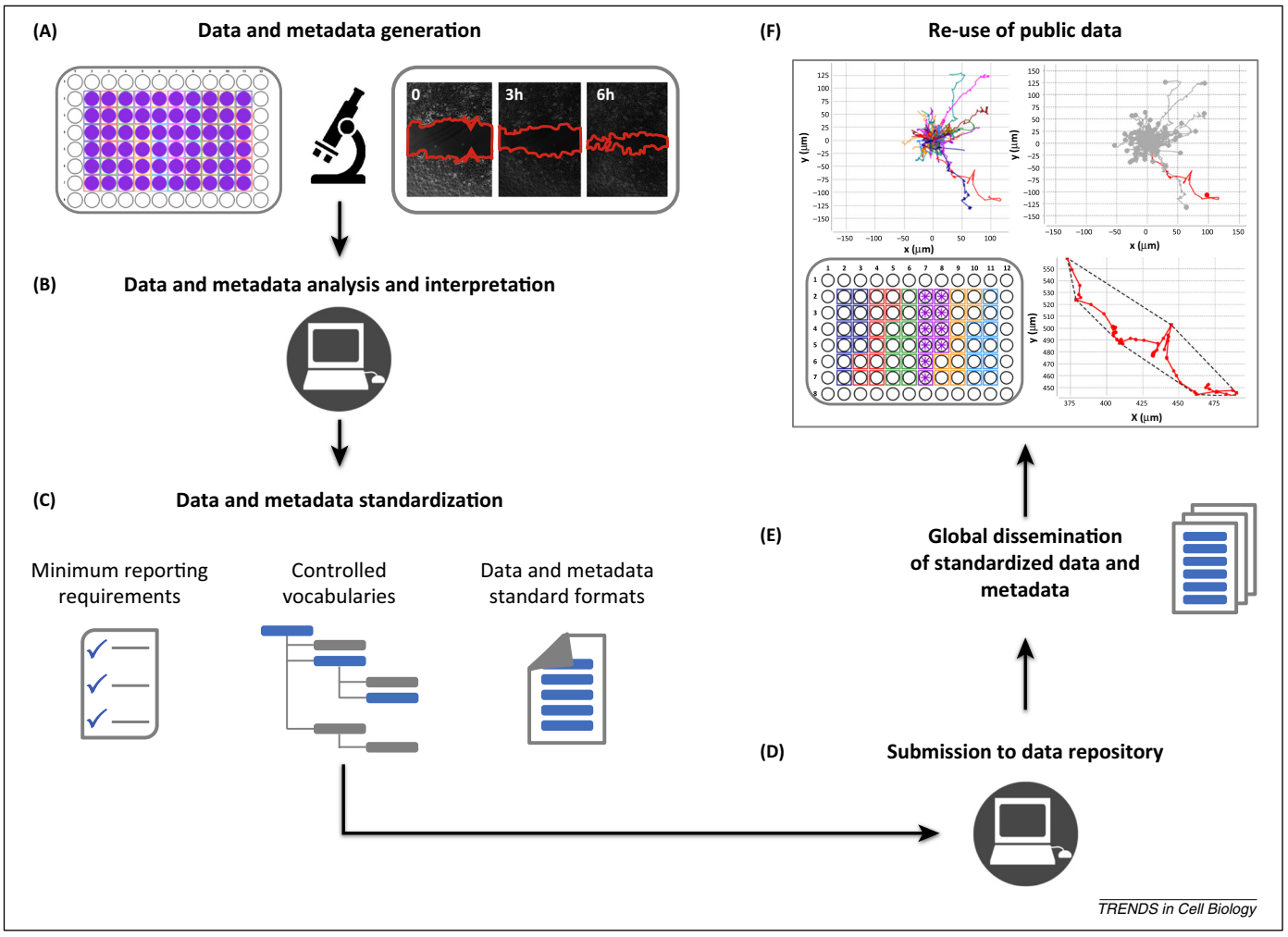

Figure 1. An example of an experimental workflow in the open data ecosystem. **(A)** Data and metadata associated with an experiment are generated. **(B)** Software is used to analyse and interpret the resulting data and associated metadata. **(C)** The collected data are formatted and reported in the relevant standards to enable data and metadata reproduction, verification, and exchange: minimum reporting requirements specify the core information to be supplied through the software tool; controlled vocabularies (CVs) are used to unambiguously annotate such units of information; and the data are exported using data and metadata standard formats. A fully standards compliant cell migration data set is ready for **(D)** submission to, and **(E)** subsequent dissemination from, a global data repository. **(F)** The open data sharing ecosystem will enable the re-use of public cell migration data, including multiscale and meta-scale analyses across large scale experiments, ultimately unlocking new knowledge in the field.

methodological context. Community-wide minimum reporting requirements have, therefore, been created in many fields, for example, for proteomics [9] and, of direct interest to cell migration, for cell perturbation experiments (http://miaca.sourceforge.net/). The global harmonization of such field-specific minimum information checklists is pursued by the BioSharing project (http://biosharing.org/).

The existing requirements can serve as a starting point to build a specific checklist for *in vitro* cell migration experiments. A tentative example of what such a list could look like is shown in Table 1: example information is provided about experimental modules and submodules, from sample preparation over image acquisition and analysis, to downstream data analysis, and laboratory metadata. A second iteration can then extend this to *in vivo* studies, which will be more challenging.

Controlled vocabularies

Minimum reporting requirements specify which information should be reported, but not yet how this information should be conveyed. The use of a common terminology thus becomes important, typically taking the form of a controlled vocabulary (CV). Again, proteomics provides an example of such a CV for the unique and unambiguous, yet detailed semantic annotation of (meta-)data [10]. Existing CVs that can be reused for cell migration experiments include the Cell Ontology [11] and the Cellular Microscopy Phenotype Ontology (http://www.ebi.ac.uk/cmpo/).

Standard data and metadata formats

When minimum reporting requirements are coupled to CVs, data and metadata can be conveyed in an unambiguous and well documented form. However, one more element is needed for successful standardization: the adoption of standard data formats. As in any data rich field, software tools are continuously applied in cell migration research to process and analyse data. However, such software can only read data presented in known formats, usually dictated by instrument vendors, and therefore implying that data can only be read by other researchers if they have access to the same instrument. Moreover, such proprietary data formats also suffer from data rot [12]. These issues can be resolved through community standard, open data formats, where considerable work

Table 1. A tentative example of what a minimal reporting requirements checklist might look like for an *in vitro* cell migration experiment[a]

Module	Submodule	Information	Example
Sample	Basic condition	Cell type	Dendritic cell
		Cell source	ATCC
		Cell species	Human
		Cell context	GFP reporter
	Pretreatment condition	Passages primary cells	Passage 4
		Medium	RPMI 1640
Assay	Assay	Dimensionality	2D
		Medium	DMEM
		Temperature	Room temperature
	Substrate	ECM type	Matrigel
		ECM concentration	2.5 mg/ml
		Post staining	YES
	Perturbation	Type	Compound
		Concentration	10 μg
Image acquisition	Time lapse	Dimensionality	2D
		Imaging modality	Bright field
		Interval	10 min
		Duration	36 h
Image analysis	Software	Software name	In-house software
	Algorithms	Segmentation	Watershed
		Tracking	Contour
Data analysis	Software	Software name	In-house software
	parameters	Biological replicates	6
		Technical replicates	3
		Readouts	Single cells tracks
		Statistics	Mann–Whitney U test
Laboratory	Experiment	User	PM
		Date	12 September 2014
		Purpose	Actin KO migration

[a]Abbreviations: DMEM, Dulbecco's modified eagle medium; ECM, extracellular matrix; GFP, green fluorescent protein; KO, knockout; RPMI, Roswell Park Memorial Institute medium.

has already been performed by Open Microscopy Environment (OME) software (http://www.openmicroscopy.org/). OME has developed widely used bioimage informatics solutions, including the OME–TIFF format that could be extended for cell migration data.

Experimental data, however, must always be accompanied by, and interpreted in the context of, overall experimental design. The existing ISA formats offer an extensible, hierarchical structure for the representation of such top-level study metadata [13], a concept that can certainly be re-used in cell migration.

Global dissemination of standardized data and metadata

Data sharing is central to scientific progress, and is fast becoming a requirement for funding or publication. Funders increasingly require grantees to share their data to maximize their value, while scientific journals require dissemination to ensure reproducibility of published results [14]. It is, therefore, logical that the centrepiece of our proposed data ecosystem should be a public repository for cell migration data. The first attempt at creating such a repository was the Cell Migration Gateway (CMG; http://www.cellmigration.org), built by the Cell Migration Consortium. Designed to be a gene-centric collection of experimental data around proteins and complexes involved in cell migration, it can be used as a starting point for the creation of a broader, more comprehensive, and future-proof cell migration data repository.

This repository should be fully conversant in the community standard formats and CVs. Furthermore, the repository should assess the adherence of datasets to the minimum reporting requirements, and perform semantic validation to check whether CV terms are used out of context. However, accepting and storing data is only a small part of the role of a repository: its most relevant function is the continuous dissemination of information. The repository thus has to offer multiple modes of access, as different users will need different types of access (e.g., manually versus automatically). It must also provide cross-references to databases in associated domains. Ideally, the repository should even serve users outside the field, enabling integrative analyses across domains in the life sciences. Over time, the system could be extended to host free software tools that execute data processing workflows, perform data analysis, and allow results interpretation.

Re-using public data: the need for novel multiscale and meta-scale analysis approaches

The data sets generated in cell migration research currently remain isolated due to the lack of a data sharing ecosystem. However, once such an ecosystem is created, it will become possible to compare and integrate data sets, and perform multiscale and meta-scale analyses across

experiments. Given the volume and the complexity of these data, however, conventional data analysis techniques will no longer be appropriate, necessitating the development of novel algorithms and approaches. These algorithms could extract features describing cell migration, to learn migratory patterns that allow classification of data sets into higher-order classes. Furthermore, such features could be used to build disease-specific models of pathogen detection, wound healing or cancer metastasis. Other algorithms could serve as automated data and metadata quality assessment tools for key data set properties [15]. Biologists and image processing experts could collaborate on small improvements in specific bioassays that can eliminate the need for novel software (e.g., colour labelling of cells migrating under high density conditions for improved tracking).

Concluding remarks

We have presented a strategy to create an open data ecosystem for cell migration research, supported by three key aspects: (i) standards and minimal reporting requirements; (ii) a public, centralized data repository; and (iii) novel analysis approaches to maximize the utility of the collected data. This ecosystem will facilitate the management, dissemination and exchange of cell migration data, allowing these data to connect to other data ecosystems in the life sciences.

Many efforts already exist towards the establishment of this ecosystem. The crucial step will, therefore, be the high-level coordination of such efforts from all interested parties – experimentalists, bioinformaticians, instrument and software vendors, funding agencies, and journals – achieved through the creation of a synergistic consortium composed of all relevant stakeholders.

Acknowledgments

P.M. and L.M. acknowledge funding from Ghent University, Ghent, Belgium, and from VIB, Ghent, Belgium.

Contributing authors

The full list of authors and affiliations is as follows:
Christophe Ampe[2], Kurt I. Anderson[4], Joseph Barry[5], Olivier De Wever[2], Olivier Debeir[6], Christine Decaestecker[6], Helmut Dolznig[7], Peter Friedl[8,9], Cedric Gaggioli[10], Benjamin Geiger[11], Ilya G. Goldberg[12], Elias Horn[13], Rick Horwitz[14], Zvi Kam[11], Sylvia E. Le Dévédec[15], Danijela Matic Vignjevic[16], Josh Moore[17,18], Jean-Christophe Olivo-Marin[19], Erik Sahai[20], Susanna A. Sansone[21], Victoria Sanz-Moreno[22], Staffan Strömblad[23], Jason Swedlow[18], Johannes Textor[24], Marleen Van Troys[2], and Roman Zantl[13]
[4] University of Glasgow, Glasgow, Scotland, UK
[5] EMBL, Heidelberg, Germany
[6] Université Libre de Bruxelles, Brussels, Belgium
[7] Medical University of Vienna, Vienna, Austria
[8] Radboud University Nijmegen, Nijmegen, The Netherlands
[9] The University of Texas, TX, USA
[10] University of Nice, Nice, France
[11] Weizmann Institute of Science, Rehovot, Israel
[12] National Institutes of Health, Bethesda, MD, USA
[13] ibidi GmbH, Munich, Germany
[14] University of Virginia, VA, USA
[15] Leiden Academic Centre for Drug Research, Leiden, The Netherlands
[16] Institut Curie, Paris, France
[17] Glencoe Software, Dundee, Scotland, UK
[18] University of Dundee, Dundee, Scotland, UK
[19] Institut Pasteur, Paris, France
[20] London Research Institute, London, UK
[21] University of Oxford, Oxford, UK
[22] King's College London, London, UK
[23] Karolinska Institutet, Solna, Sweden
[24] Utrecht University, Utrecht, The Netherlands.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.tcb.2014.11.005.

References

1 Friedl, P. and Gilmour, D. (2009) Collective cell migration in morphogenesis, regeneration and cancer. *Nat. Rev. Mol. Cell Biol.* 10, 445–457
2 Kramer, N. *et al.* (2013) *In vitro* cell migration and invasion assays. *Mutat. Res.* 752, 10–24
3 Friedl, P. and Bröcker, E.B. (2000) The biology of cell locomotion within three-dimensional extracellular matrix. *Cell. Mol. Life Sci.* 57, 41–64
4 Alexander, S. *et al.* (2013) Preclinical intravital microscopy of the tumour-stroma interface: invasion, metastasis, and therapy response. *Curr. Opin. Cell Biol.* 25, 659–671
5 Hulkower, K.I. and Herber, R.L. (2011) Cell migration and invasion assays as tools for drug discovery. *Pharmaceutics* 3, 107–124
6 Lock, J.G. and Strömblad, S. (2010) Systems microscopy: an emerging strategy for the life sciences. *Exp. Cell Res.* 316, 1438–1444
7 Meijering, E. *et al.* (2012) Methods for cell and particle tracking. *Methods Enzymol.* 504, 183–200
8 Edelman, L.B. *et al.* (2010) *In silico* models of cancer. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 2, 438–459
9 Taylor, C.F. *et al.* (2007) The minimum information about a proteomics experiment (MIAPE). *Nat. Biotechnol.* 25, 887–893
10 Mayer, G. *et al.* (2013) The HUPO proteomics standards initiative – mass spectrometry controlled vocabulary. *Database (Oxford).* 2013, bat009
11 Bard, J. *et al.* (2005) An ontology for cell types. *Genome Biol.* 6, R21
12 Martens, L. *et al.* (2005) Do we want our data raw? Including binary mass spectrometry data in public proteomics data repositories. *Proteomics* 5, 3501–3505
13 Sansone, S-A. *et al.* (2012) Toward interoperable bioscience data. *Nat. Genet.* 44, 121–126
14 Anon (2007) Democratizing proteomics data. *Nat. Biotechnol.* 25, 262
15 Foster, J.M. *et al.* (2011) *A posteriori* quality control for the curation and reuse of public proteomics data. *Proteomics* 11, 2182–2194