SCIENTIFIC DATA

HELPING YOU PUBLISH, DISCOVER AND REUSE RESEARCH DATA

Calling for submissions in Fall 2013, launching in Spring 2014 **nature.com/scientificdata**

Dr Ruth Wilson
Nature Publishing Group

Dr Susanna-Assunta Sansone
University of Oxford e-Research Centre

Dr Andrew Hufton
Nature Publishing Group

CREDIT
Credit for depositing & sharing your data

REUSE
Complete, curated & standardized descriptions enable the reuse of your data

QUALITY
Rigorous community-based peer review

DISCOVER
Helping you find datasets relevant to your research

OPEN
Available to all through a Creative Commons license

SERVICE
Rapid peer review & publication

WHY *SCIENTIFIC DATA*?

Researchers, funders, community standards initiatives and learned societies alike are calling for new ways to make scientific data more available and reusable.

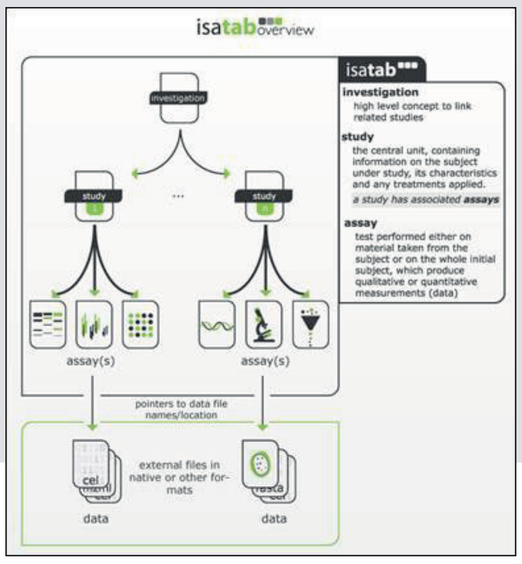
Scientific Data aims to meet these needs through the publication of Data Descriptors. These describe scientifically valuable datasets and provide credit to authors who share their data and add meaningful information around that data to ensure it is interpretable and reusable.

WHAT IS *SCIENTIFIC DATA*?

- *Scientific Data* is a peer reviewed Open Access, online-only publication containing Data Descriptors that describe and explain datasets, supported by an APC model.
- Focus is on data interpretability and reuse and not analysis and hypothesis.
- Peer review is optimised to evaluate data quality and reuse.
- Research data files will be stored in one or more public, community-recognized repositories.

STRUCTURED CONTENT

- Data Descriptors are complemented by machine readable structured information to further promote reuse which will be released as CC0.
- Curated and standardised by in-house curator.



DATA DESCRIPTORS

complements both journal articles and repository records

OPEN

SUBJECT CATEGORIES
» Induced pluripotent stem cells
» Proteomic analysis
» Microarray analysis

Proteomic profiles of human embryonic stem cells, induced-pluripotent stem cells and precursor fibroblasts

Javier Munoz[1] and Albert J.R. Heck[2,3]

Assessing relevant molecular differences between human-induced pluripotent stem cells (hiPSCs) and human embryonic stem cells (hESCs) is important, given that such differences may impact their potential therapeutic use...

STRUCTURED SUMMARY

Design Type(s)	cell type comparison design • growth condition intervention design
Measurement Type(s)	protein expression profiling • transcription profiling assay
Technology Type(s)	mass spectrometry assay • DNA microarray
Factor Type(s)	cell line • growth condition
Sample Characteristic(s)	*Homo sapiens* • embryonic stem cell line • embryonic fibroblast cell line • foreskin fibroblast cell line

Background & Summary
Human embryonic stem cells (hESCs) are capable of self-renewal and multi-lineage differentiation (that is, pluripotency)[1] ...

Methods
Cell Culture
Human induced pluripotent stem cell lines IMR90_iPS and 4Skin_iPS were obtained from James A. Thomson (Wisconsin National Primate Research Center, University of Wisconsin-Madison, Madison) who ...

METHODS
Detailed reproducible methods describing processes or experiments that generated the data outputs. Enough detail is provided to allow other researchers to interpret and repeat (if required) the study.

DATA RECORDS
A description of each data record associated with a Data Descriptor, including the repository where this information is stored and an overview of the data files and their formats.

Data Records
All the samples used in this study are summarized in Table 1. Consistent identifiers are used in Tables 2 and 3 to allow mapping between the proteomic and transcriptomic data outputs.

Data Record 1
The raw data, peaklists (.mgf), ProteomeDiscoverer result files (.msf) and ProteomeDiscoverer workflow files (.xml) have been uploaded to ProteomeXchange (http://www.proteomexchange.org/) with the following accession number PXD000134 (ref. 67; Table 2).

TECHNICAL VALIDATION
Any experiments or analyses that are needed to support the technical quality of the dataset.

Technical Validation
Confirmation of pluripotency and experimental design
Two experiments using MSbased proteomic experiments using two different hiPS cell lines were conducted (Fig. 1). In Experiment 1, IMR90_iPS were compared to hESCs (HES3) and to the parental cell line, IMR90_Fibro. In Experiment 2 ...

Usage Notes
The data is publicly available in different formats including Thermo LTQ Orbitrap (e.g..raw) files and Mascot Generic Format (e.g..mgf) peaklists. Originally[40], the data was...

USAGE NOTES
Instructions or advice that may help other researchers reuse the dataset.

TABLES
Detailed information accounting for the sample or subject in the study, and the data-generating assays applied to each sample – plus the resulting data outputs.

DATA CITATIONS
Included in references section.

DOES NOT INCLUDE
- In depth analysis or tests of hypotheses
- New scientific conclusions
- Exploratory analysis (e.g. clustering)

This is a sample Data Descriptor derived from:
The quantitative proteomes of human-induced pluripotent stem cells and embryonic
Munoz, J, Low T. Y., Kok, Y.J., Chin, A., Frese, C. K., Ding, V., Choo, A. & Heck A.J.
Mol. Syst. Biol. 7, 550 (2011)

Supported by: figshare credit for all your research · isatools COLLECT | CURATE | ANALYSE | SHARE | PUBLISH · biosharing · DRYAD · nature publishing group npg