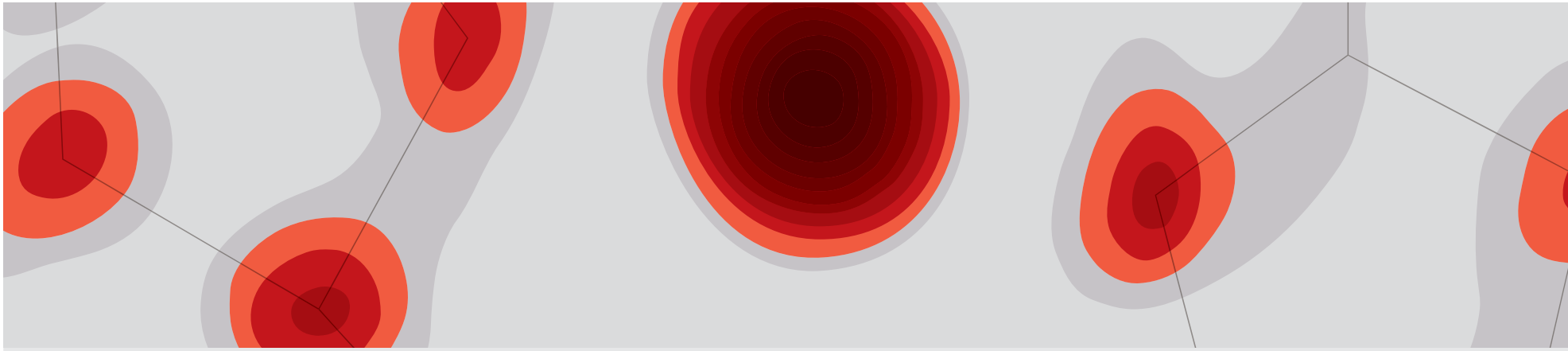
Part of SPRINGER NATURE



HRA Webinar: Intro to Data Sharing

Andrew L. Hufton

Managing Editor, *Scientific Data*

https://www.nature.com/sdata/

nature research

Science has always required showing your evidence

1

Backing up your claims with evidence has always been a key part of science

Roger Bacon, 1214 – 1292

Opus Majus

“ Theories supplied by reason should be verified by sensory data, aided by instruments, and corroborated by trustworthy witnesses

Opus Tertium

“ The strongest argument proves nothing so long as the conclusions are not verified by experience.

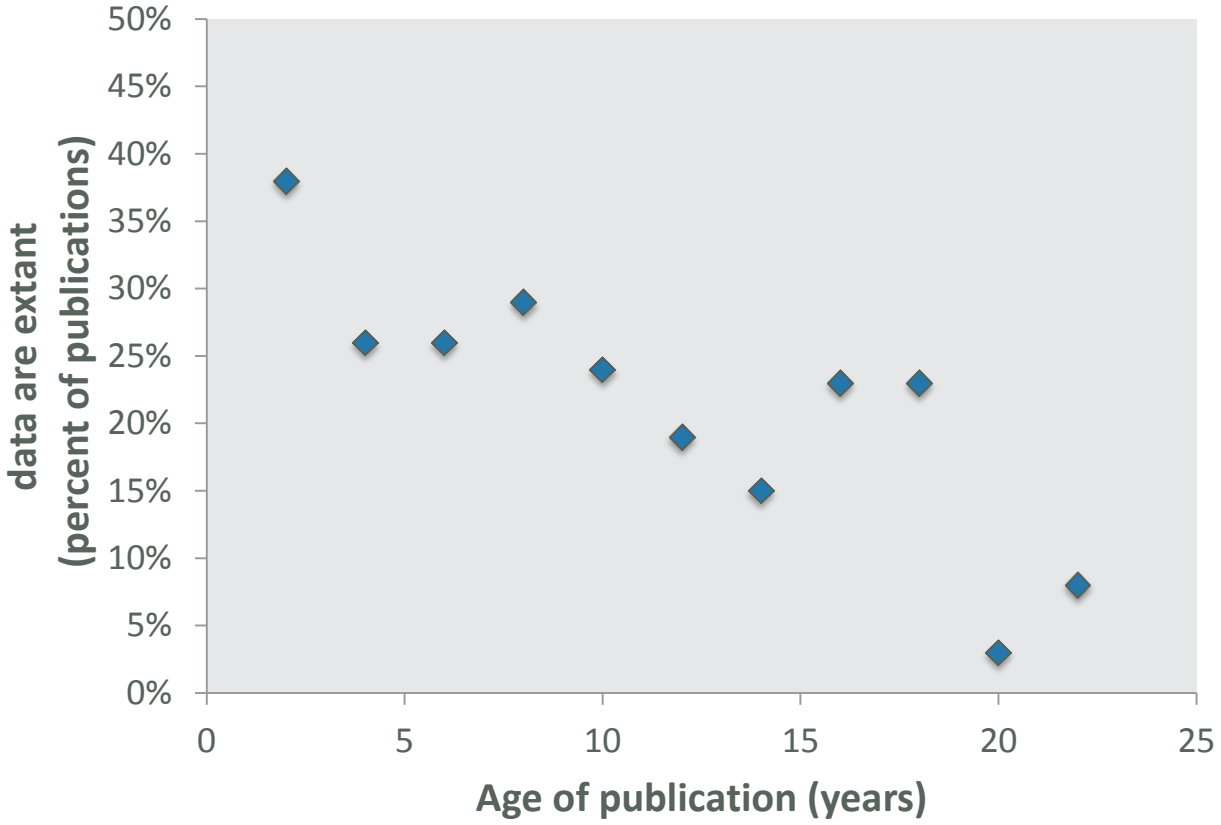
Fundamental sharing policy for *Nature* and the Nature research journals

> An inherent principle of publication is that others should be able to replicate and build upon the authors' published claims. A condition of publication in a Nature journal is that **authors are required to make materials, data, code, and associated protocols promptly available** to readers without undue qualifications. Any restrictions on the availability of materials or information must be disclosed to the editors … [and] … in the submitted manuscript.

Supporting data must be made available to editors and peer-reviewers at the time of submission for the purposes of evaluating the manuscript.

See http://www.nature.com/authors/policies/availability.html

Sharing upon request has problems



Replotted from:
Vines *et al*. *Current Biology* (2014)
doi:10.1016/j.cub.2013.11.014

Raw data at *Dryad*
doi: 10.5061/dryad.q3g37

When developing a data sharing policy…

2

Think about how you can make the data useful to others…

Open data is about more than disclosure – it must be "FAIR"

- Findable
- Accessible
- Interoperable
- Re-usable

Wilkinson *et al. Sci. Data* **doi:10.1038/sdata.2016.18** (2016)
https://www.nature.com/articles/sdata201618

Think about why you are sharing data…

- Support data sharing within defined collaborations (i.e. sharing with friends)
- Help others critically evaluate and reproduce an authors' claims (i.e. sharing with critics)
- Allow others to use data in separate research projects, including overlapping or competitive research (i.e. sharing with competitors)

Publish your data

3

the data paper

SCIENTIFIC DATA

- A clear, peer reviewed description of data, to maximize usage
- Citable publications that give credit for reusable data
- Visit our journal homepage to learn more: https://www.nature.com/sdata/

Get Credit for Sharing Your Data

Publications will be indexed and citeable.

Open-access

Data Descriptors are published under a Creative Commons Attribution licence (CC BY). Each publication supported by CC0 metadata.

Focused on Data Reuse

All the information others need to reuse the data; no interpretative analysis, or hypothesis testing

Peer-reviewed

Rigorous peer-review focused on technical data quality and reuse value

Promoting Community Data Repositories

Not a new data repository; data stored in community data repositories

Find the right repository for your data

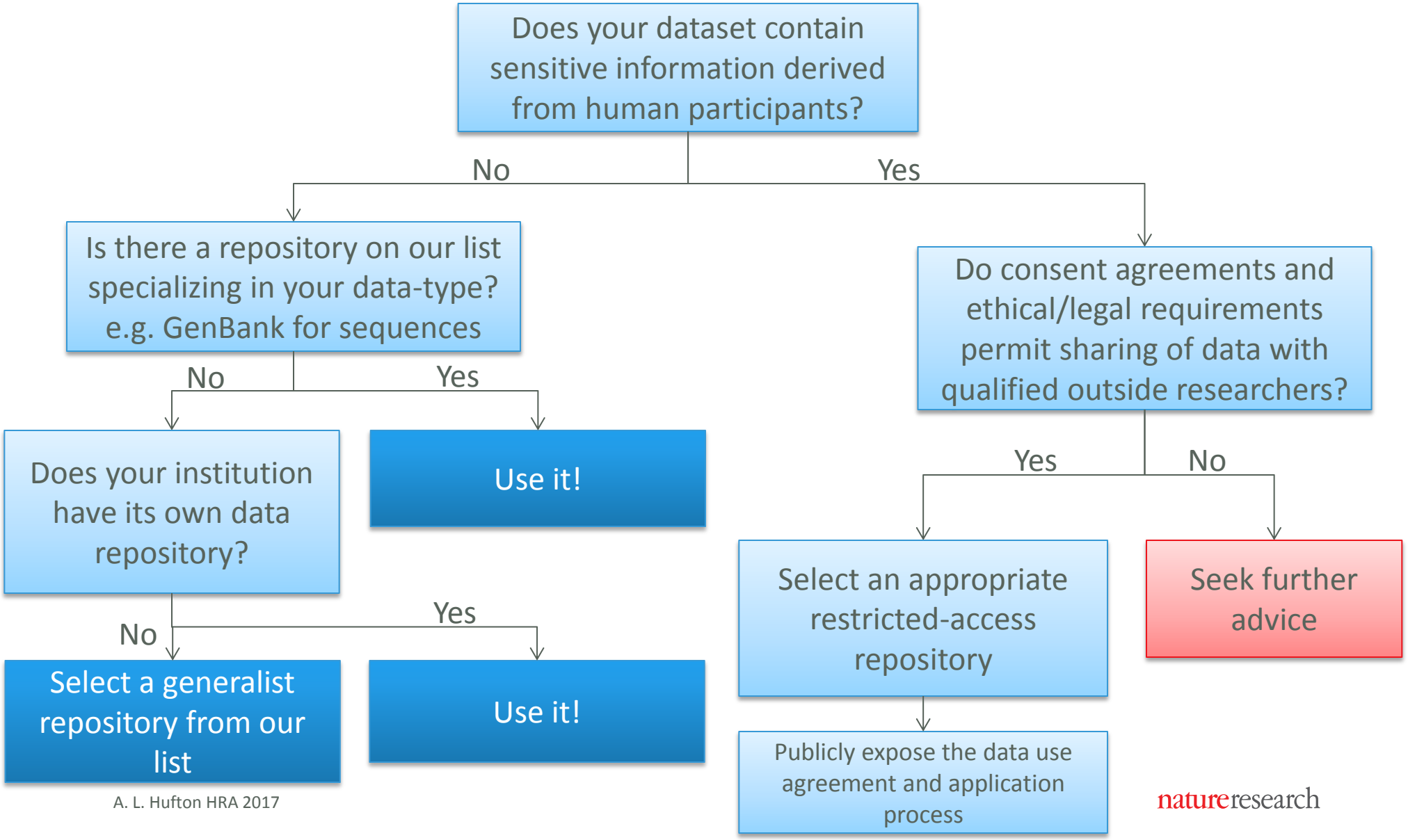
3

General features to look for in a data repository

- Quality curation
- A commitment to long-term preservation
- Features that support collaborative analysis
- Features that allow you keep data private until you are ready to publish.
- Open data licensing options (e.g. CC0 or CC BY)

An overly simple guide to picking the right data repository

https://www.nature.com/sdata/policies/repositories



A. L. Hufton HRA 2017

nature research

Repositories that can handle sensitive human data

Always seek approval and guidance from your institutional ethical authority and/or data access committee before depositing human data to a third-party repository. In most cases, data should be carefully anonymized or pseudonymized before deposition.

Genotype-phenotype data archives

- dbGAP (http://www.ncbi.nlm.nih.gov/gap)
- EGA (http://www.ebi.ac.uk/ega/)

National disease-specific databases

- National Addiction & HIV Data Archive Program (http://www.icpsr.umich.edu/icpsrweb/NAHDAP/)
- National Database for Autism Research (http://ndar.nih.gov/)

Social science databases

Extensive experience with human-derived datasets, often can handle diverse kinds of clinical and health-related data

- UK Data Service ReShare (http://reshare.ukdataservice.ac.uk/)
- openICPSR (http://www.openicpsr.org/)

Generalist repositories

figshare

- In-browser data viewers, make tables and code easily previewable
- Media files immediately playable
- No link to a peer-reviewed publication required
- 100 GB of free storage available via *Scientific Data*, data kept private during peer-review
- Unlimited public storage available via figshare.com for researchers willing to make data immediately public

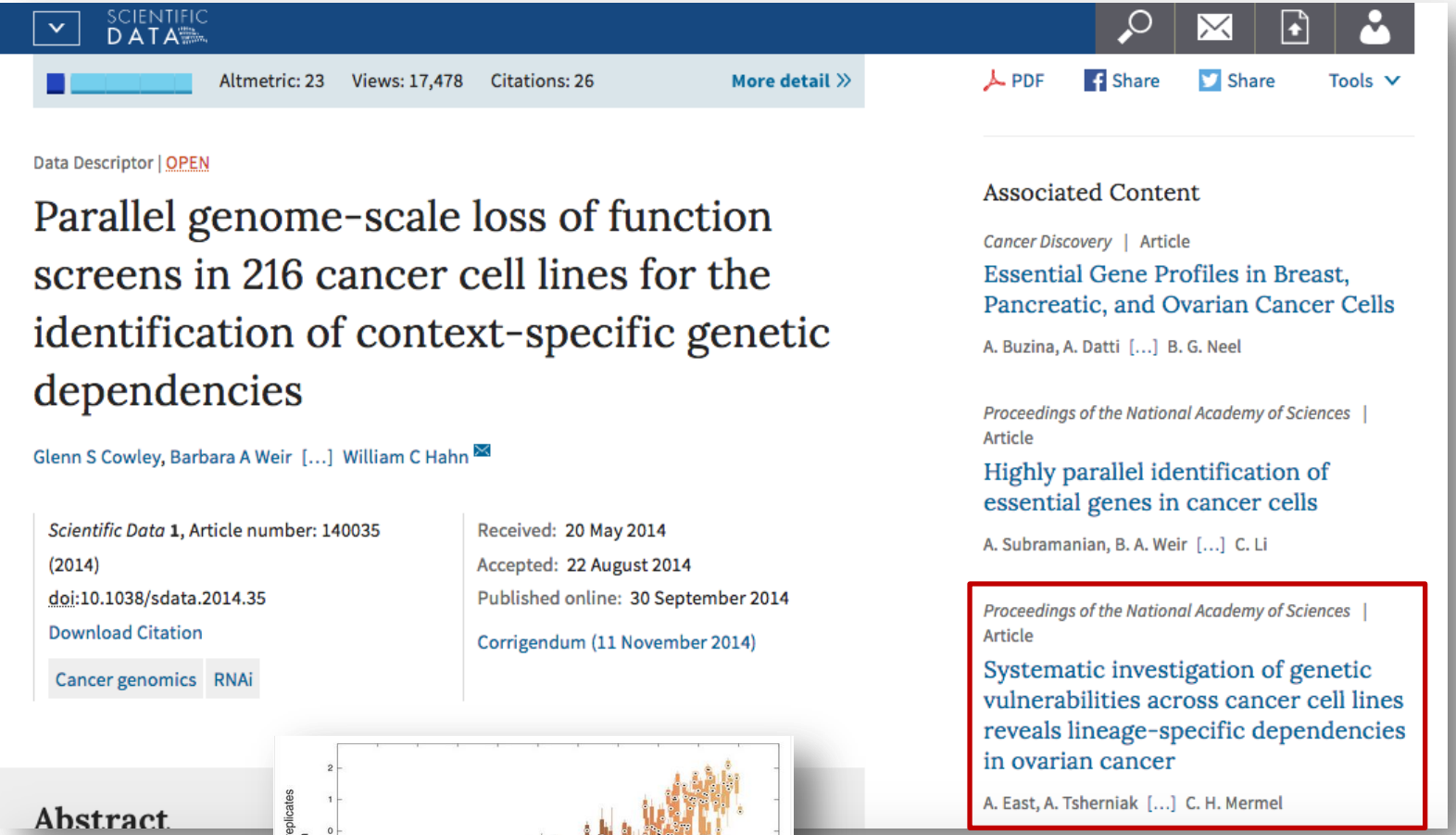
Dryad Digital Repository (http://datadryad.org/)

- $120 USD for first 20 GB, and $50 USD for each additional 10 GB
- Curation support helps authors with basic file naming & upload checking, and screens for inappropriate human data
- Only accepts data associated with a specific publication

Others: Harvard Dataverse (http://dataverse.harvard.edu/), Open Science Framework (http://osf.io/), Zenodo (http://zenodo.org/)

Data reuse is powerful

SCIENTIFIC DATA



SCIENTIFIC DATA

Altmetric: 23 Views: 17,478 Citations: 26 More detail »

PDF Share Share Tools

Data Descriptor | OPEN

Parallel genome-scale loss of function screens in 216 cancer cell lines for the identification of context-specific genetic dependencies

Glenn S Cowley, Barbara A Weir […] William C Hahn

Scientific Data **1**, Article number: 140035 (2014)
doi:10.1038/sdata.2014.35
Download Citation

Cancer genomics RNAi

Received: 20 May 2014
Accepted: 22 August 2014
Published online: 30 September 2014

Corrigendum (11 November 2014)

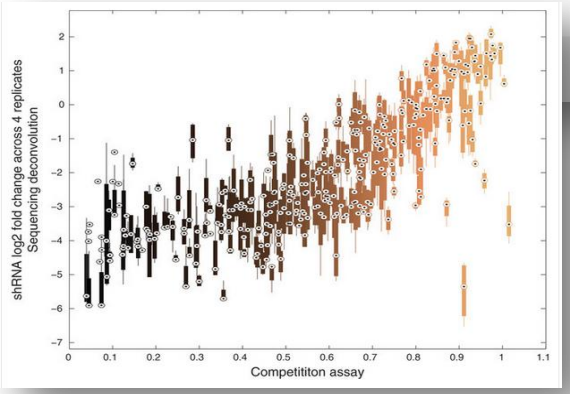
Abstract

Associated Content

Cancer Discovery | Article
Essential Gene Profiles in Breast, Pancreatic, and Ovarian Cancer Cells
A. Buzina, A. Datti […] B. G. Neel

Proceedings of the National Academy of Sciences | Article
Highly parallel identification of essential genes in cancer cells
A. Subramanian, B. A. Weir […] C. Li

Proceedings of the National Academy of Sciences | Article
Systematic investigation of genetic vulnerabilities across cancer cell lines reveals lineage-specific dependencies in ovarian cancer
A. East, A. Tsherniak […] C. H. Mermel



- **Screen results and in-depth analysis published in 2011 at *PNAS***
- **Full screen data published at *Scientific Data* in 2014**
- Data at **figshare**
- Data Descriptor cited 94 times according to Google Scholar!

nature research

Get the most from your data

Preserve it
Encourage reuse
Get credit

Encourage others to do the same

SCIENTIFIC DATA

Thanks!

Managing Editor
Andrew L. Hufton
andrew.hufton@nature.com

Honorary Academic Editor
Susanna-Assunta Sansone

Data Curation Editor
Varsha Khodiyar

Senior Publishing Assistant
Joseph Salter

Head of Data Publishing
Iain Hrynaszkiewicz

Visit nature.com/scientificdata

Email scientificdata@nature.com

Tweet @ScientificData

Supported by DRYAD biosharing figshare credit for all your research isatools COLLECT | CURATE | ANALYSE | SHARE | PUBLISH

nature research