Scientific Record Keeping and Data Management

Xiaokun Shu
Department of Pharmaceutical Chemistry
Cardiovascular Research Institute
April 2015

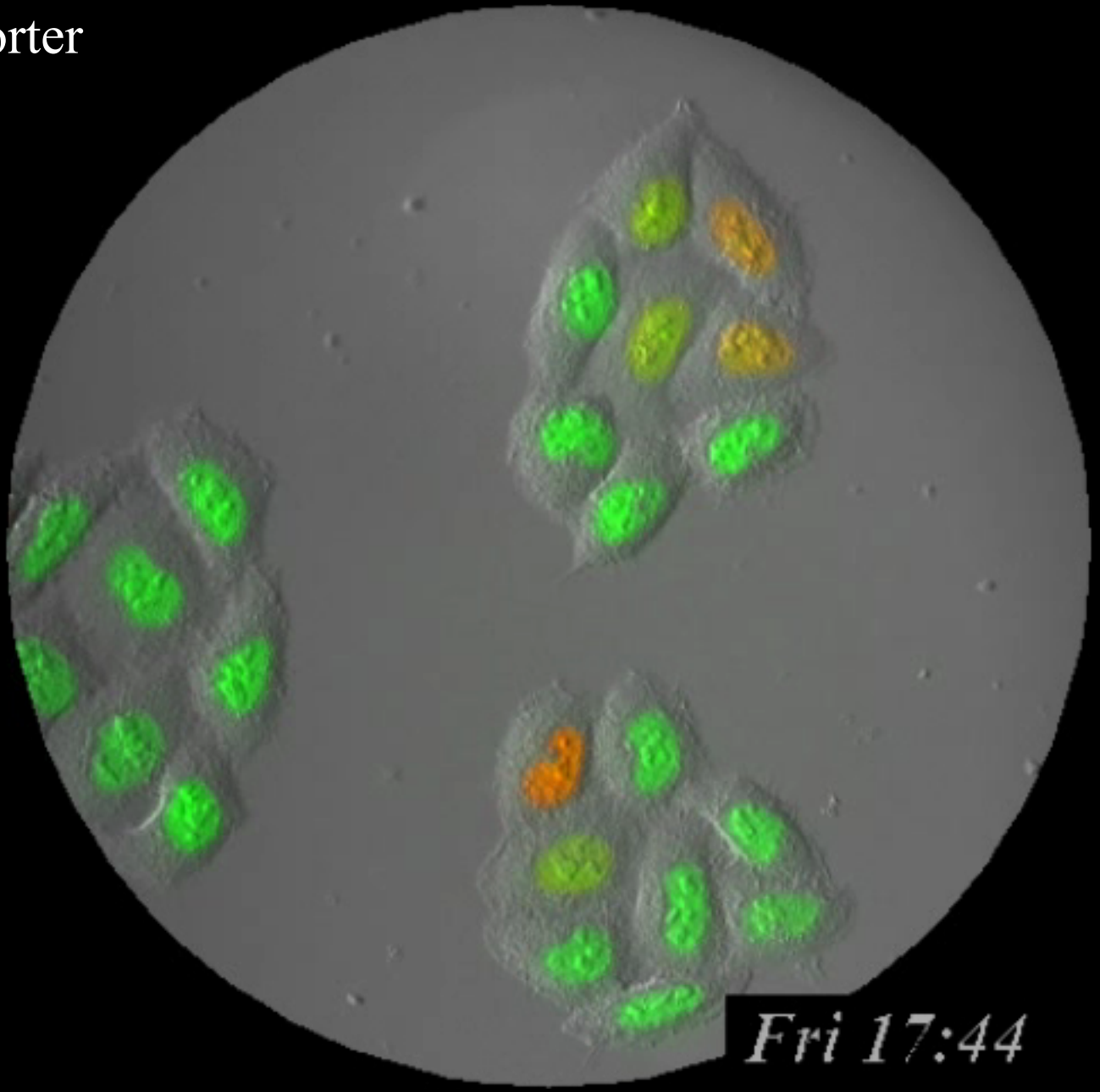
Modified from Laura Goodwin, UCSF Career Center

Data - The Scientific Keystone

Publications	Collaboration	Funding
Animals/Human	Research Data	Mentoring
Intellectual Property	Industry	Public Integrity

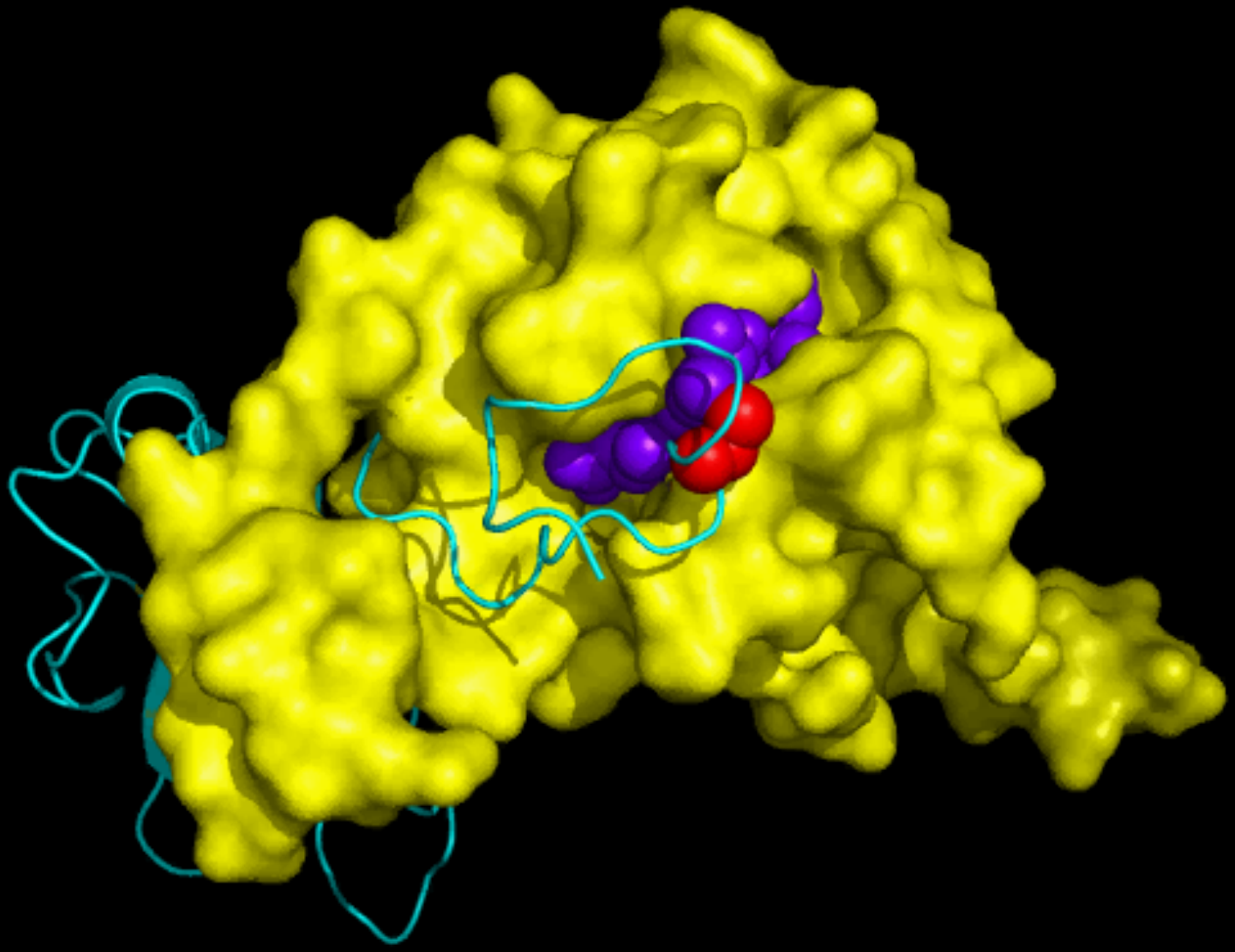
- Definition of research data
- Distinctions between ownership, control, and access to data
- Standards for keeping laboratory notebooks
- Standards for maintaining data
- Data retention and destruction of data
- Standards for mentoring new investigators regarding data
- Standards for sharing data
- Standards for presenting data in publications and grants

Cell cycle reporter



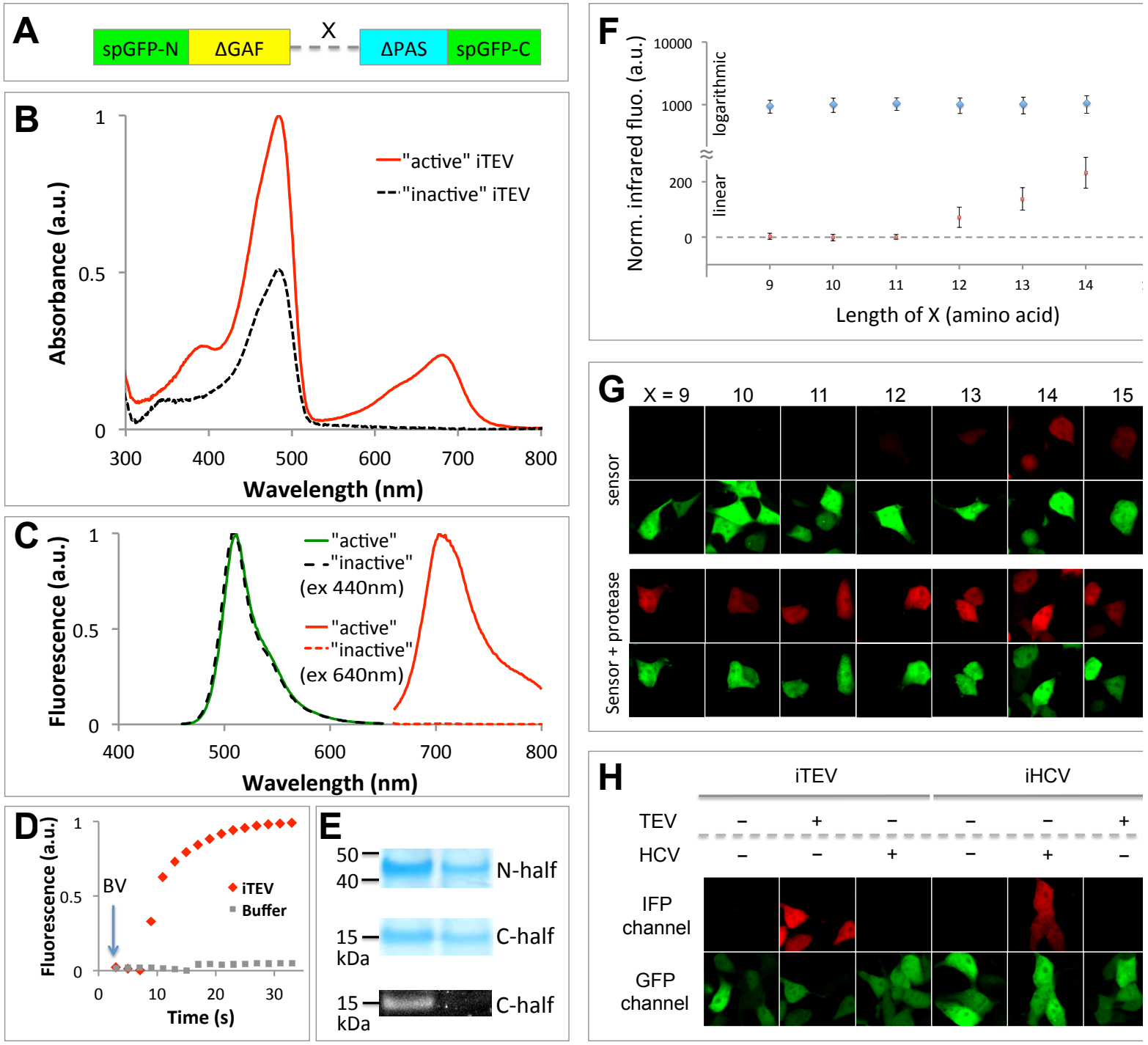
Cell 132, 487 (2008)

Rational design of a fluorogenic protease reporter

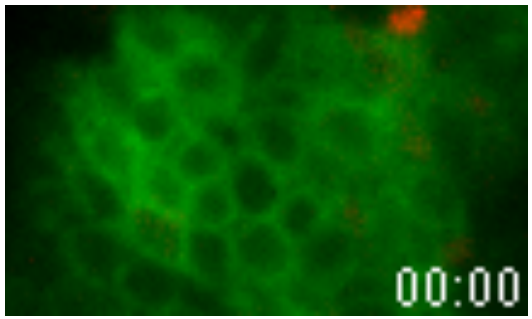


To et al PNAS 112, 3341 (2015)

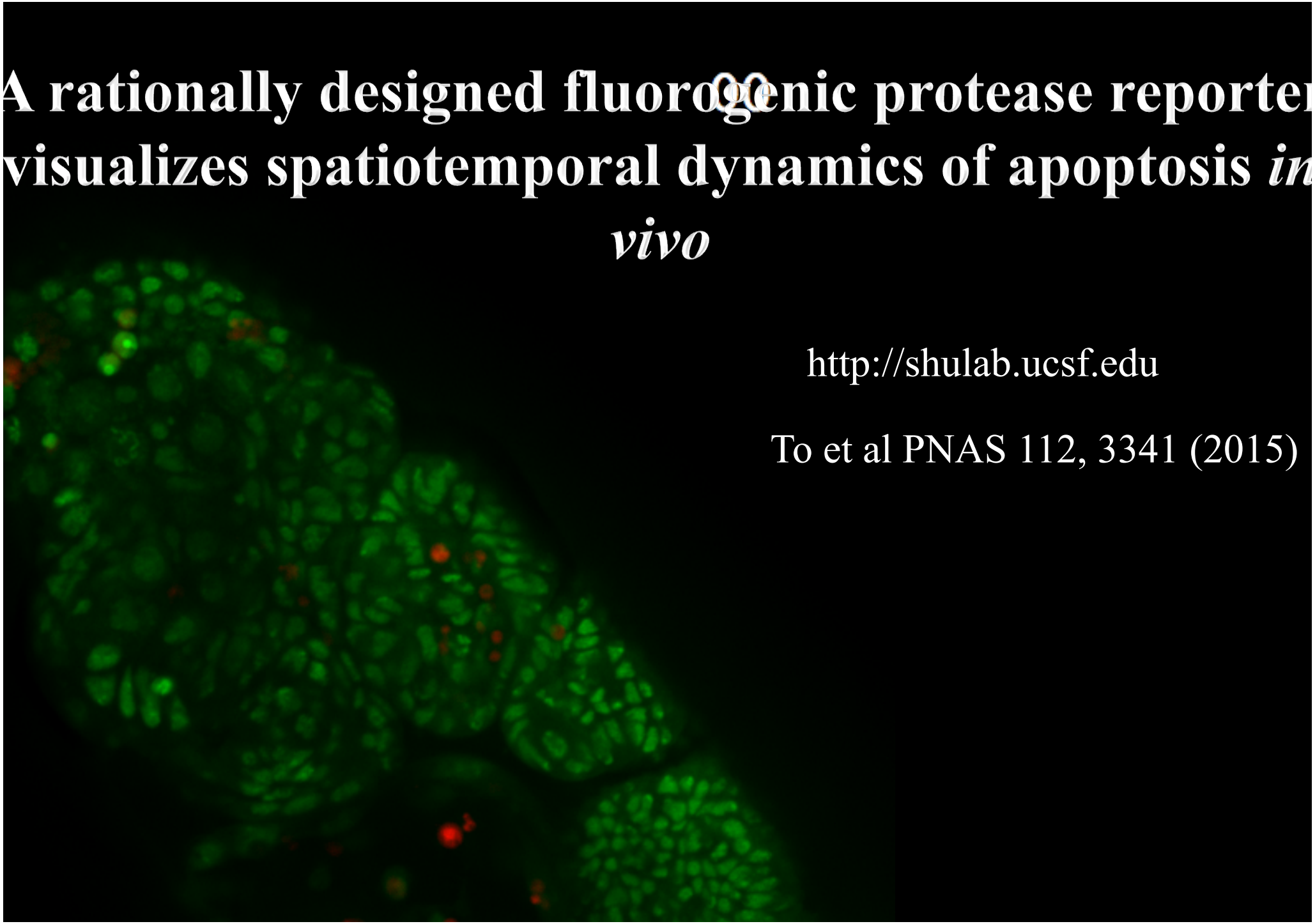
Data collection, analysis and presentation



nfrared fluorescent caspase reporter visualizes cell apoptosis



A rationally designed fluorogenic protease reporter visualizes spatiotemporal dynamics of apoptosis *in vivo*

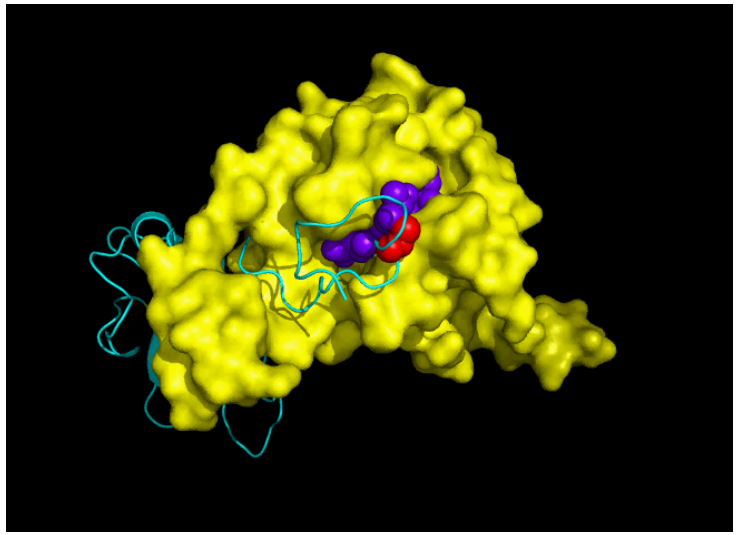


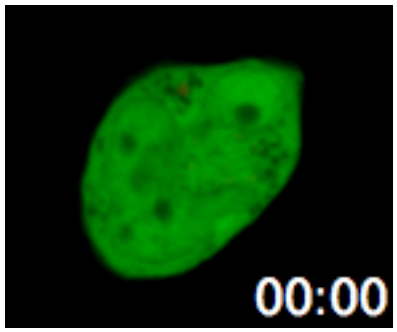
http://shulab.ucsf.edu

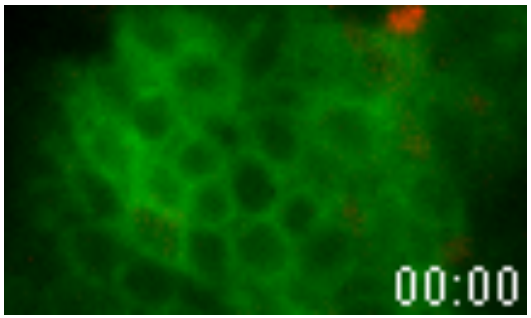
To et al PNAS 112, 3341 (2015)

A rationally designed fluorogenic protease reporter visualizes spatiotemporal dynamics of apoptosis *in vivo*

- Design of a protease reporter
- Data collection, analysis, presentation
- Patent application
- Publication
- Grant application
- Data sharing
- Collaboration







Definition of Data

The Federal Acquisition Regulation (45 CFR 27.401) defines data as "recorded information regardless of form or the media on which it may be recorded. The term includes technical data and computer software."

Problems in Research Record-Keeping

- Lack of formal training for students - the "apprenticeship model"
- Increasingly entrepreneurial nature of biomedical research
- Growing litigiousness of the research community and critical importance of reliable data
- Growing complexity in kinds of research data that must be recorded and managed.

Photographs

Computer data storage

Print outs

Video and audio tape

Biologics

Data Ownership, Control, and Access

- Ownership of Data

  "Both as a matter of law and practical neccessity, institutions should assert the ownership of research data."

  (Final lecture. Office of Technology Management)

- Control of Data

  "Institutions should explicitly delegate control of research data to the laboratory director or principal investigator, except in extraordinary circumstances. The PI should have authority and responsibility to control, safeguard, and determine access to data.

- Access to Data

  Standards for data sharing

Ownership of data: Industry

- Company
  - Owns all data and is also the custodian
  - Holds all patents
  - Holds all copyright
- PI is an employee
  - Limited rights

Ownership: Role of Lab Member?

- Original lab notebooks
  - Kept with PI
  - Student may be able to copy lab books upon leaving lab
    - If data public, student can copy
    - If not public, student and PI should confer with Office of Technology Management because untimely release could jeopardize patent rights

Standards for Keeping Lab Notebooks

neurosurgery.medschool.ucsf.edu/neurosurgery_research/guidelines.html

"The Hard Copy of Choice"

- Good notebook that lasts long
- Permanently bound, sewn binding
- Pages consecutively numbered
- Good quality paper
- Permanent ink – not pencil, not unusual colored inks

A Viable Alternative

Loose-leaf binders may be used at the PI's discretion:

- All pages numbered consecutively prior to use
- Experiment identification numbers must be present and follow a system developed in the lab
- The system must be described on the first page of each binder

The Notebook - Contents

- Table of Contents page(s) including all experiments.
- For each experiment:

Title of study• PI's name• Date the study starts; date it ends• Associate investigator's name(s) (i.e. staff research associate/ postdoctoral fellow/resident/graduate student)• Brief statement of hypothesis or study goals• Cell line (passage no.)• Animal strain and supplier• Specific animal identification no. (large animals)• Source of analyzed material• Tumor type & passage no.• Drug type (lot no. and/or source), dose(s), dilution(s)• Radiation source and dose• Special reagents (e.g. antibodies, probes)• Cell culture batch/medium• Serum batch/medium• Experimental design• Study-specific treatment groups, projected number of subjects, and all other elements of the study design with reference to specific techniques or protocols from the methodology notebook [refer to each specific protocol or technique by its designation (e.g. number) in the methodology notebook].• Statistical methodology added to or deviating from that in the methodology notebook.• A 'time line' illustrating the sequence of study events (e.g. start--cells added--medium added--etc.)• Specific notes about special procedures or steps that differ from the techniques specified in the methodology notebook (e.g. changes in incubation time or temperature, concentration of trypsin, tumor cell inoculum, infusion rates, anesthetic procedure). Any variance from the routine procedures recorded in the methodology notebook should be noted in detail [refer to each specific routine protocol or technique by its designation (e.g. number) in the methodology notebook].• Raw data, or explicit instructions for locating the raw data or retrieving them from storage (see 3.4.4). Of particular importance are notations about excluded data or animals, with detailed information about why those data or animals were excluded.• A brief conclusion of the experiment, including a 'value judgment' about the validity of the experiment, whether the study needs to be repeated to validate it, what can and cannot be definitively concluded from the data, and other observations. Simple concluding descriptions such as 'bad study,' or 'data suspect,' are not acceptable. It is essential to document why the study or data were considered suspect. Justifications for positive judgments similarly should be recorded.

Raw Data

Now, where did I put those gels?

- Pasted in or enclosed in plastic sleeves is fine, but must travel with the experiment
- "Data too unwieldy to include are listed in the experimental notebook as they are collected, are described sufficiently for recognition, and are annotated with the name of investigator and explicitly where the data can be found (e.g. location of the tape or disc and its identification number)"
- Physical location of the computer, name, directory of the file (drive/your name/year/month/date/data)

Recording Data - Standards

- Experiments should be replicatable by someone not familiar with your system.
- Consecutive numbering and recording, always.
- Errors crossed out but legible.
  - No erasures
  - No liquid paper
  - For error in experimental entry – cross out, make a new entry, initial and EXPLAIN
  - Don't remove or add pages
  - If large section is in error, put X through it, write "OMIT" and put new entry adjacent with explanation

Recording Data: Patent Protection

- For patent purposes
  - Have witness review and sign pages each week
  - Corroborating witness
    - Not directly involved in the same project
      - Not supervisor, not co-inventor
    - Reads and understands the work
    - Writes “read and understood”, signs and dates
  - Keep work uninterrupted or document why
    - “On vacation”
  - Avoid interpretation
    - just the facts

Retention and Destruction of Data

- "PIs store all data notebooks and related data and records in their laboratory for 5 years after the date when funding for a study ends. They may then continue to store them in their laboratory or may make arrangements to have them moved to the University's storage facility; both the PI and the research administrator keep a record of the information necessary to retrieve the materials from the facility. Data notebooks and related data and records for any study may not be destroyed."

- FDA - 10 year retention for non-clinical studies

- Patent Office - 23 years after the issue of a patent for original data supporting patent claims

Standards for Sharing Data

- Required by the NIH

  - For any grant over $500,000
  - Data must be shared no later than the date of publication acceptance
  - A plan for data sharing is included for each grant

- Data to be shared:

  "The research data necessary to validate research findings"

- Data that cannot be requested:

  " Lab notebooks, manuscript drafts, analyses, communications with collegues, etc."

The Methodology Notebook

A brief word

There are practices and standards for the keeping of a notebook in the lab that describes commonly used methods and can be referred to by individual experiments.

- Numbered, dated, documented as to when each technique goes into and out of use.

Case Study 1: A Change of Protocol

ituation

lua is doing a postdoctoral fellowship in a laboratory that studies cancer treatment. In the experiment he is overseeing, a cancer-prone strain of mice is allowed to develop visible tumors and then receives xperimental drugs to observe the effects on the tumors. Hua notices that the tumors are interfering wi he ability of some of the mice to eat and drink. She also notices that some of the mice are weaker and nore emaciated than the others, which she suspects is a consequence of their feeding difficulties. The rotocol for the experiment states that the mice will be treated only if they exhibit obvious signs of pai r discomfort.

When she mentions her concerns to another postdoctoral fellow, he suggests not raising the issue with he rest of the lab. The mice will be euthanized as soon as the experiment is over, and their nutritional tatus probably has little or no effect on the drug treatment. Furthermore, if it proved necessary to hange the experimental protocol, the previous work would be invalidated and the Institutional Animal Care and Use Committee would need to be notified.

Questions:

. What can Hua do to get more information about the issue?
. If she decides to raise the issue with others, what is the best way to do so?
. Should the original protocol have been approved?

urce: On Being a Scientist: A Guide to Responsible Conduct in Research, 3rd edition, Washington, D.C.: National Academy Press (2009), page 14.

Collection of Data

- Use proper controls
  - Control for all confounders
  - Use controls established by other labs
  - Or randomize
- Use proper sample sizes
  - Consult with statistician beforehand
- Collect all relevant data from experiment
  - Not just those that support the hypothesis
- Be open to events without bias or assumptions
- Keep data for suitable length of time
  - Often, data "lost" and unable to be verified

www.unmc.edu/ethics/data/data_int.htm

Selection of Data: Rationale

- Aim is to observe signal, exclude noise

- Unethical selection is a major temptation
  - Avoid selection based on bias
  - Avoid selection based on preconceived expectations
  - Avoid selecting best data to make a better figure

Standards for Presenting Data

#1 - Fraud

Journals are beginning to implement standards (ex: JCB):

"No specific feature within an image may be enhanced, obscured, moved, removed, or introduced. The grouping of images from different parts of the same gel, or from different gels, fields, or exposures must be made explicit by the arrangement of the figure (e.g., using dividing lines) and in the text of the figure legend. Adjustments of brightness, contrast, or color balance are acceptable if they are applied to the whole image and as long as they do not obscure or eliminate any information present in the original. Nonlinear adjustments (e.g., changes to gamma settings) must be disclosed in the figure legend."

How Serious is the Problem?

- The managing editor of JBC estimated in April of 2005 that roughly 20% of ACCEPTED manuscripts contain at least one figure that must be remade due to inappropriate image manipulation.

- ORI Cases Involving Questioned Data:

‘89-90	2.5%
‘95-96	5.7%
‘99-00	14.3%
‘01-02	~30%

(John Krueger, “Forensic Examination of Questioned Scientific Images”)

The Take-Home

When in doubt, Document!

Case Study 2: Discovering an Error

Situation:

Two young faculty members—Marie, an epidemiologist in the medical school, and Yuan, a statistician in the mathematics department—have published two well-received papers about the spread of infections in populations. As Yuan is working on the simulation he has created to model infections, he realizes that a coding error has led to incorrect results that were published in the two papers. He sees, with great relief, that correcting the error does not change the average time it takes for an infection to spread. But the correct model exhibits greater uncertainty in its results, making predictions about the spread of an infection less definite.

When he discusses the problem with Marie, she argues against sending corrections to the journals where the two earlier articles were published. "Both papers will be seen as suspect if we do that, and the changes don't affect the main conclusions in the papers anyway," she says. Their next paper will contain results based on the corrected model, and Yuan can post the corrected model on his Web page.

Questions:

1. What obligations do the authors owe their professional colleagues to correct the published record?
2. How should their decisions be affected by how the model is being used by others?
3. What other options exist beyond publishing a formal correction?

urce: On Being a Scientist: A Guide to Responsible Conduct in Research, 3rd edition, Washington, D.C.: National Academy Press (2009), page 14.