Guidelines to annotate experimental protocols

SCOPE

We are manually annotating experimental protocols in life sciences. We want to identify words or phrases that can be related to: i) the **Sample(s)** tested in a protocol, ii) **Instruments** used, iii) **Reagents** employed, and the overall iv) **Objective** of a protocol –**SIRO** elements. Before reading this document please look at the slides illustrating how to use the tool.

These **four** elements are common across protocols in life sciences. The manual identification of these elements will help us to: i) enrich our controlled vocabularies and, ii) facilitate information retrieval.

WHAT SHOULD BE ANNOTATED?

Words or phrases related to:

- Sample(s), specimen(s) or organism(s) to be tested.
- Instruments (including software) and consumables.
- Reagents, chemical compounds, solutions or mixtures used.
- Objective or purpose of the protocol.

SOME EXAMPLES…

The **sample** tested in a protocol may be an organism or a part of it. Some examples include:

SAMPLE	Whole organism	Scientific name: *Arabidopsis thaliana*, *Oriza sativa*, *mangifera indica, Mus musculus.* Common name: Mousear Cress, rice, mango, mouse.
	Anatomical part	leaf, stem, cells, tissues, membranes, organs, skeletal system, muscular system, nervous system, reproductive system, cardiovascular system, etc.
	Biomolecules	Nucleic acids: Deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). Proteins: enzymes, structural or support proteins (keratin, elastin, collagen), antibodies, hormones, etc.
	Body fluids	Blood serum, saliva, semen, amniotic fluid, cerebrospinal fluid, gastric acid, etc.

The **instruments** used in protocols include high-throughput equipment, software and consumables. Some examples:

INSTRUMENTS	High-throughput equipment	Liquid Handling Platforms, Real-Time PCR Detection System, Microplate Reader, etc.
	Instruments	Goggles, Bunsen burner, spot plate, pipet, forceps, test tube rack, mortar and pestle, etc.
	Laboratory glassware	Beaker, Erlenmeyer flask, graduated cylinder,

		volumetric flask, etc.
	Standard equipment	Balances, shakers, centrifuges, refrigerators, incubators, thermocyclers, fume hood, etc.
	Consumables	Weighing dishes, pipette tips, gloves, syringes, petri dishes, test tubes, micro centrifuge tubes, glass slides, filter paper, etc.

The **reagents** used in protocols include buffers, solutions, culture media and kits. Some examples:

REAGENTS	Chemical compound/ Substance	Glucose, ethanol, glycerol, chloroform, acetic acid, isopropyl alcohol, etc.
	Solutions / buffers	70% ethanol, 10X PCR buffer, phenol:chloroform:isoamyl alcohol, etc.
	Kits	Nucleic acid purification kits, virus purification kits, PCR screening kits, etc.
	Cell culture media	Nutrient media, minimal media, selective media, differential media, etc.

The **objective** of a protocol is a formal statement describing the goal; what we do want to achieve by executing the protocol? An example:

Part of the speech	Source
"Here we present a detailed protocol for Smart-seq2 that allows the generation of full-length cDNA and sequencing libraries by using standard reagents"	doi:10.1038/nprot.2014.006

ANNOTATION PROCESS

The annotators should carry out the following steps in this specific order:

1. **Read the whole document**. Read the document from start to end, making no annotations to get an understanding of the processes described in the protocol.

2. **Mark the entities**. Read the document a second time and annotate it. The process is simple:

   a. detect the word or phrase of interest,

   b. highlight by selecting the test of interest with the mouse, then

   c. use one of the labels (tags) that you will see on the right hand side of the screen. The tags are sample, reagent, instrument, objective

This annotation task is just about attaching a label (sample, instrument, reagent or objective) to a word or phrase. Examples:

- "Yeast" → sample
- "70% Ethanol" → reagent
- "PCR apparatus" → instrument
- "We present a protocol for isolating DNA from ancient bones" → objective

d. **Mark each occurrence of an entity**. For example if the word "ethanol" appears in different parts of the document, make sure to mark each occurrence. Please, bear in mind that sometimes samples, equipment and reagents are not listed in the materials sections.

e. **Reduce the noise in the annotation**. Avoid excess. Don't mark an entire paragraph that includes more than one entity associated to the same label (see row 3 in the table below). Also, avoid making incomplete annotations (see the first row of the table).

✓ Good annotations	✗ Bad annotations
9| Add glycerol to 15% or DMSO to 7% (vol/vol),	9| Add glycerol to 15% or DMSO to 7% (vol/vol)
This protocol was developed for single cells but works equally well on purified total RNA of a few picograms or more. A flow-	This protocol was developed for single cells but works equally well on purified total RNA of a few picograms or more. A flow-

2.1 Annotating

2.1.1 Sample/Specimen/Organism

In the Nature Protocol (doi:10.1038/nprot.2007.427), the authors use several names for the same sample. Some of them are:

1. *"collection of gene-deletion mutants in Saccharomyces cerevisiae"*
2. *"yeast deletion mutants"*

Note: the annotators should annotate the different names given to the sample tested.

Let's focus on *"collection of gene-deletion mutants in Saccharomyces cerevisiae"*, this could be annotated as follows and both are correct:

- *"collection of gene-deletion mutants in Saccharomyces cerevisiae"* OR
- *"collection of gene-deletion mutants in Saccharomyces cerevisiae"*

Please, annotate other potential samples that could be used. For instance, in doi:10.1038/nprot.2007.427:

> *"…the general method of studying pooled samples with barcode arrays can also be adapted for use with other types of samples, such as mutant collections in other organisms, short interfering RNA vectors and molecular inversion probes."*

In this text there are 3 samples that could be tested:

1. *mutant collections in other organisms,*
2. *short interfering RNA vectors,*
3. *molecular inversion probes.*

Therefore, you should add three annotations, each one tagged as Sample/Organism, as follows:

> *"…the general method of studying pooled samples with barcode arrays can also be adapted for use with other types of samples, such as mutant collections in other organisms, short interfering RNA vectors and molecular inversion probes."*

A **bad annotation practice**, is highlighting a part of speech that includes the 3 samples and add a unique tag, for example:

> *"…the general method of studying pooled samples with barcode arrays can also be adapted for use with other types of samples, such as mutant collections in other organisms, short interfering RNA vectors and molecular inversion probes."* *

* Note: in our database, each annotation is an entry. If the text s wrongly annotated, then the database of annotations will consider that the protocol "x" has only one sample/Organism entry.

2.1.2 Instruments

In protocols (such as doi:10.1038/nprot.2007.427), it is sometimes mentioned the capacity of the containers used. Some examples are:

- 96-well microliter plates
- 250 ml flasks
- 0.5 ml microfuge tubes

When this information is available, the annotation should include the storage capacity of the containers used and described in the protocol.

Please, avoid highlighting the entire list of instruments/equipment for adding a unique tag; the database of annotations will consider that the protocol "x" has only one Instrument entry.

Examples of good a bad practices annotating instruments:

Good annotation of instruments	Bad annotation of instruments
(❖) • 48-well plates (Greiner, part no. 677102) • 250 ml culture flasks • 0.5 ml microfuge tubes suitable for boiling (Eppendorf 0.5-ml Safe-Lock microcentrifuge tubes; Sigma, cat. no. T8911)	• 48-well plates (Greiner, part no. 677102) • 250 ml culture flasks • 0.5 ml microfuge tubes suitable for boiling (Eppendorf 0.5-ml Safe-Lock microcentrifuge tubes; Sigma, cat. no. T8911)
Equipment (❖) • Aerodisc 0.2 mm filters (Pall Life Sciences, cat. no. 4192) • Syringes (Becton Dickinson, cat. no. 309653) • Nunc Omni trays (VWR, cat. no. 62409-600) • 96-well pin tool (V&P Scientific, cat. no. VP407A)	**Equipment (⌘)** • Aerodisc 0.2 mm filters (Pall Life Sciences, cat. no. 4192) • Syringes (Becton Dickinson, cat. no. 309653) • Nunc Omni trays (VWR, cat. no. 62409-600) • 96-well pin tool (V&P Scientific, cat. no. VP407A)

❖ each line or bullet point is annotated as an instrument.
⌘ the full list was highlighted and annotated as an instrument

2.1.3 Reagents

The protocols should include a list of the chemical and biochemical compounds and solutions (e.g. buffers, Cell culture media) used in a protocol. Sometimes, some chemical compounds are used as a solute in a solution. Please annotate the different concentration grades of the reagents (e.g. G418, 1,000X G418 stock (200 mg ml^{-1}), 1,000X G418 stock –these are highlighted in blue in the example below). Illustrating how to annotate see the following example from doi:10.1038/nprot.2007.427. Other reagents, also listed in the example, are highlighted in green

"REAGENTS

.
.

G418 (Agri-Bio, cat. no. 3000)

.
.

REAGENT SETUP

1,000X G418 stock (200 mg ml^{-1}) *Dissolve 5 g of G418 in 25 ml of dH2O. Filter-sterilize using a 0.2 mm filter and a syringe. Shield from light by wrapping bottle in foil. Store at 4 ºC.*

YPD + 200 µg ml^{-1} G418 rectangular plates *Mix 10 g of yeast extract, 20 g of peptone, 20 g of dextrose, 20 g of agar and 1 liter of dH2O to a 2-liter flask with a stir bar. Autoclave. Allow media to cool to approximately 50 ºC with gentle stirring. Add 1 ml of 1,000X G418 stock. Stir gently for an additional 1 min to ensure that the drug is evenly mixed. Pour into Nunc Omni trays, 50 ml per tray; sufficient for approximately 20 plates. Store at 4 ºC.*

YPD liquid + 200 lg ml $^{-1}$ G418 *Mix 10 g of yeast extract, 20 g of peptone, 20 g of dextrose and 1 liter of dH2O to a 1-liter bottle. Autoclave. Allow media to cool to approximately 50 ºC. Add 1 ml of 1,000X G418 stock. Store at 4 ºC.*

Up primer mix *Dissolve uptag and Buptagkanmx4 each in dH2O at 100 pmol µl $^{-1}$ and mix in a 1:1 ratio. Note that the uptag oligo is also used in mixed oligonucleotides, so take care to leave enough for use in both mixes. Store at 20 ºC."*

If you want to include the quantity used for reagent, it is also fine. For instance:

- 20 g of peptone ***instead of*** 20 g of peptone
- 1 liter of dH2O ***instead of*** 1 liter of dH2O

2.1.4 Objective

The objective or goal of the protocol in most cases is described in the abstract section. Please, highlight only the part of the text that describes the objective of the protocol. For instance,

"… from a single pooled culture. Here, we present protocols for the study of pooled cultures of tagged yeast deletion mutants with a tag microarray. This process involves..."

3. **Add a comment indicating an annotation or decision that was hard to make**. Comment in the annotation tool; the tool has a comment field above the labels. Please use it; this will be useful for us in order to understand and respond to those difficulties.

4. **Time to solve doubts.** Planning personalized sessions (via skype, slack, email, etc.) to solve doubts is simply a matter of contacting me (oxgiraldo@gmail.com). These sessions help us to know, for example:

    - any questions or uncertainty that you may have about the annotations and guidelines,
    - anything unclear or ambiguous in the guideline,
    - things that you consider important and that were not covered by the labels for the annotations in this task.
    - Bugs in the annotation tool and/or issues.

5. **Notify the finalization of the annotation task**. Each annotator should notify via email the finalization of his/her annotation task.